NASA CR-141895
ERIM 109600-18-F

Final Report

MULTISPECTRAL PROCESSING BASED ON GROUPS OF RESOLUTION ELEMENTS

WYMAN RICHARDSON AND JAMES M. GLEASON
Infrared and Optics Division

MAY 1975

Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Earth Observations Division
Houston, Texas 77058
Contract No. NAS9-14123, Task V
Technical Monitor: Dr. A. Potter/TF3

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
BOX 618 • ANN ARBOR • MICHIGAN 48107

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. NASA CR- ERIM 109600-18-F	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle MULTISPECTRAL PROCESSING BASED ON GROUPS OF RESOLUTION ELEMENTS		5. Report Date May 1975 6. Performing Organization Code
7. Author(s) Wyman Richardson and James M. Gleason		8. Performing Organization Report No. 109600-18-F
9. Performing Organization Name and Address Environmental Research Institute of Michigan Infrared and Optics Division P.O. Box 618 Ann Arbor, Michigan 48107		10. Work Unit No. Task V 11. Contract or Grant No. NAS9-14123
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Johnson Space Center Houston, Texas 77058		13. Type of Report and Period Covered Final Technical Report May 15, 1974 through March 14, 1975 14. Sponsoring Agency Code

15. Supplementary Notes

The work was performed for the Earth Observations Division. Dr. Andrew Potter (TF3) was the technical monitor.

16. Abstract

The two subject of this report are nine-point classification and boundary detection. A nine-point classification rule is one that makes a decision for each pixel on the basis of data from that pixel and its eight immediate neighbors. Three previously-studied nine-point rules had performed substantially better than the usual one-point rule on field interiors but had clumsily represented boundary areas. Three new nine-point rules are defined and in preliminary tests compared with the previously studied rules. One of the new rules performed well in boundary areas but with reduced efficiency in field interiors. Another did just the opposite. A third combined best performance on field interiors with good sensitivity to boundary detail.

The basic thresholded gradient and some modifications of it, have been investigated as a means of boundary point detection. The test results of these methods were for the most part unsatisfactory.

The hypothesis testing methods of closed boundary formation developed by LARS, Purdue, have also been tested and evaluated. Some reasonable results for satellite data were less satisfactory.

A thorough and sophisticated analysis of the boundary detection problem has been initiated. Statistical signal detection and parameter estimation techniques are being employed to analyze various formulations of the problem. These formulations will permit the atmospheric and sensor system effects on the data to be thoroughly analyzed. Various boundary features and necessary assumptions can also be investigated in this manner. The results of the one basic formulation which was investigated are encouraging.

17. Key Words	18. Distribution Statement
Remote Sensing Multispectral Processing Classification Rules Multispectral Recognition Spatial-Discrimination Boundary Detection	Initial distribution is listed at the end of this document

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	120	

NOTICES

Sponsorship. The work reported herein was conducted by the Environment Research Institute of Michigan for the National Aeronautics and Space Administration, Earth Observations Division, Lyndon B. Johnson Space Center, Houston, under contract NAS9-14123. Dr. Andrew E. Potter/TF3 is Technical Monitor for NASA. Contracts and grants to the Institute for the support of sponsored research are administered through the Office of Contracts Administration.

Disclaimers. This report was prepared as an account of Government-sponsored work. Neither the United States, nor the National Aeronautics and Space Administration (NASA), nor any person acting on behalf of NASA:

(A) Makes any warranty or representation, expressed or implied with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method or process disclosed in this report may not infringe privately owned rights: or

(B) Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used above, "person acting on behalf of NASA" includes any employee or contractor of NASA, or employee of such contractor, to the extent that such employee or contractor of NASA or employee of such contractor prepares, disseminates, or provides access to any information pursuant to his employment or contract with NASA, or his employment with such contractor.

Availability Notice. Requests for copies of this report should be referred to:

National Aeronautics and Space Administration
Scientific and Technical Information Facility
P.O. Box 33
College Park, Maryland 20740

Final Disposition. After this document has served its purpose, it may be destroyed. Please do not return it to the Environmental Research Institute of Michigan.

PREFACE

This report describes part of a comprehensive and continuing program of research concerned with advancing the state-of-the-art in remote sensing of the environment from aircraft and satellites. The research is being carried out for NASA's Lyndon B. Johnson Space Center, Houston, Texas, by the Environmental Research Institute of Michigan (ERIM), formerly the Willow Run Laboratories of The University of Michigan. The basic objective of this multidisciplinary program is to develop remote sensing as a practical tool to provide the planner and decision-maker with extensive information quickly and economically.

Timely information obtained by remote sensing can be important to such people as the farmer, the city planner, the conservationist, and others concerned with problems such as crop yield and disease, urban land studies and development, water pollution, and forest management. The scope of our program includes (1) extending the understanding of basic processes; (2) discovering new applications, developing advanced remote-sensing systems, and improving automatic data processing to extract information in a useful form; and (3) assisting in data collection, processing, analysis, and ground-truth verification.

The research described herein was performed under NASA Contract NAS9-14123, Task V and covers the period from May 15, 1974 through March 14, 1975. Andrew Potter (TF3) was the NASA Contract Technical Monitor. The program was directed by Richard R. Legault, Vice-President of ERIM and Head of the Infrared and Optics Division, Jon D. Erickson, Head of the Information and Analysis Department, and Richard F. Nalepka, Principal Investigator and Head of the Multispectral Analysis Section.

Part I of this report was written by Wyman Richardson. The author gratefully acknowledges the helpful suggestions of James M. Gleason, Richard J. Kauth, Michael J. McClary, and Robert B. Crane of ERIM. Part II was written by James M. Gleason. The author wishes to express his appreciation for the contribution of Robert B. Crane and Wyman Richardson. The ERIM number of this report is 109600-18-F.

CONTENTS

FIGURES

TABLES

1
SUMMARY

The two subjects of this report are nine-point classification and boundary detection.

Nine-point classification rules decide what material to assign to a pixel on the basis of data from that pixel and from its eight immediate neighbors. They are applicable whenever a pixel is likely to represent the same material as its neighbors, as is the case with agricultural data. The purpose of such rules is to gain recognition accuracy at only a slight extra cost in processing time.

In the previous work [1] three such rules were defined, implemented and tested.

LIKE9, the nine-point maximum likelihood rule, amounts to adding, for each material, the nine multivariate normal exponents and choosing the material with the smallest sum. To prevent occasional alien points from disturbing the decision rule, only the m smallest exponents are summed, where m is a number between 1 and 9.

AVE9, averages the nine data points and then applies the usual one-point rule QRULE. It is modified to delete the t largest and t smallest of the nine data values in each channel.

VOTE9, applied after QRULE decisions have been made on the nine pixels, assigns to the center pixel the material most frequently recognized among the nine pixels.

These three nine-point rules substantially outperformed QRULE on field interiors, but maps made by them showed distortion and loss of detail on the boundaries. They are based on the rigid premise that all nine pixels represent the same material and they performed clumsily when the premise failed.

---

[1] W. Richardson, A Study of Some Nine-Element Decision Rules, Technical Report, 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan 1974.

To reach the goal of developing nine-point rules that are accurate on boundaries, and thereby suitable for satellite data with its high frequency of boundary points, two nine-point rules were derived from assumptions less rigid than neighborhood conformity.

BAYES9 is based on the assumption that a pixel probably represents the same material as its neighbor, the degree of dependence specified by a parameter θ between 0 (independence) and 1 (complete dependence).

PRIOR9 makes a Bayesian decision on the center pixel based on prior probabilities estimated from neighborhood data values. The estimated prior probability of a material is the average, over nine pixels, of the posterior probability of that material at each pixel.

PREF9 uses as its decision criterion the estimated prior probability just defined for PRIOR9. It is conceptually an improved voting rule that takes account of all the information at each pixel rather than just a vote for the winning material.

In preliminary tests an aircraft data, BAYES9 with θ = .3, .5, .7, and .9 performed as well on field interiors as the three previously-studied rules and was considerably more sensitive to fine detail on the boundaries. For θ = .1, the field interior results were nearly as good. The maps improved in fidelity to known boundaries as θ decreased from .9 to .1.

PRIOR9 made an excellent map of boundary areas, but its improvement over QRULE in field interiors was only half that of the best nine-point rules.

PREF9 ranked with the best on field interiors - a little better than the other voting rule VOTE9 - but resembled LIKE9 and AVE9 in its clumsy representation of boundary areas.

A null test (i.e., a means of deciding "none of these") was included in each of the new rules by defining a null category of all materials not associated with a specific distribution and giving it a flat distribution of height ε. The effectiveness of this method of deciding null was verified by the good maps made by BAYES9 and PRIOR9.

A three-stage plan for testing nine-point rules on LANDSAT data from the LACIE experiment is presented. The first stage is a test of all nine-point rules on two LACIE intensive study sites. The measure of comparison will be the percentage of interior points from ground-inspected fields that are correctly classified. On the basis of the first-stage and aircraft data results, a small number of rules will be chosen for further testing. The second stage will be the same as the first, but with fewer rules and more sites. The third stage will be a comparison of second stage rules as acreage estimators.

The basic thresholded gradient and some modifications to it, have been tested and evaluated as a means of boundary point detection. The computational efficiency of those methods is particularly appealing. The modifications were developed to utilize some characteristics of the boundaries not utilized by the basic technique.

The hypothesis testing techniques developed by LARS, Purdue have been tested and evaluated as a means of closed boundary formation. Two methods have actually been developed, one using first-order statistics only and the other also using second-order statistics. These techniques were investigated to gain a better insight into the closed boundary formation problem and also to better ascertain their performance.

The performance of the gradient methods was for the most part unsatisfactory. These methods detected many true boundary points but also detected too many false boundary points. They are effective as an easily implemented means of boundary enhancement for visual examination. However, difficulties in choosing a proper threshold and a tendency to emphasize randomly occurring variations, are two major problems which all of these techniques suffer from.

The performance of the hypothesis testing techniques is more difficult to categorize. The most significant aspect of these techniques is the field building algorithm which guarantees closed boundaries. The method using first-order statistics only performed better than that using second-order statistics on the data sets which were tested. The results of both methods were significantly poorer for the satellite data set than for the aircraft data set which were tested.

After testing the gradient and hypothesis testing techniques, it was decided that the boundary detection problem must be approached in a thorough and sophisticated manner. The approach which was adopted is to formulate the boundary detection problem in increasingly complex steps and perform a vigorious mathematical analysis of each. The formulations of the problem will contain the atmospheric and sensor system effects on the data. Different boundary features and assumptions will also be investigated as part of these different formulations. Statistical signal detection and parameter estimation methods will be used for the analysis.

Only one basic formulation of the problem has been investigated. The solution procedure which was derived is easily implemented and is also optimal under certain assumptions. The analysis of this formulation is indicative of the power of this approach.

PART I

NINE-POINT CLASSIFICATION

Wyman Richardson

2

INTRODUCTION

Multispectral classification rules in current use are based on information from one pixel at a time. The objective of this investigation is to increase the accuracy of multispectral recognition by developing classification rules that use data from groups of pixels.

The research has focused on "nine-point" rules (i.e., those which take into account, when classifying a pixel, data from the eight surrounding pixels) because they offer hope of satisfying the following requirements: 1) more accurate than purely spectral rules, 2) practical with respect to execution speed and computer storage, 3) preserve as much resolution as possible, and 4) suited to agricultural surveys such as LACIE.

Three such rules were implemented in the previous contract and tested on one data set with encouraging results[1]. They are all "nine-point" rules, that is to say they use data from the 3 x 3 grid formed by the pixel in question and its eight immediate neighbors.

LIKE9, the nine-point likelihood rule, is the maximum likelihood decision rule derived from the assumption that the nine elements are an independent random sample from a multivariate normal distribution. It amounts to adding, for each material, the nine multivariate normal exponents and then choosing the material with the smallest sum. To prevent occasional alien points from disturbing the decision rule, we have modified it to sum only the m smallest exponents, where m = 1,...,9.

AVE9, the trimmed mean rule, averages the nine data points and then applies the one-point (i.e., the usual) rule. To lessen its sensitivity to alien points we have deleted the t largest and the t smallest values of the nine in each channel, where t = 0,...,4.

[1] W. Richardson, A Study of Some Nine-Element Decision Rules, Technical Report 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan 1974.

VOTE9, the voting rule, is applied after one-point decisions have been made on the nine pixels. It assigns to the center pixel the material most frequently recognized among the nine pixels. In case of a tie, the one-point decision on the center pixel is used.

To compare and rank these three rules and the usual one-point rule QRULE, we ran a quantitative test by counting the number of points misclassified within each of 42 field interiors in the Imperial Valley, California. A result of the test was the following best-to-worst ranking of rule performance: LIKE9 with $m = 9$, VOTE9, AVE9 with $t \neq 0$, AVE9 with $t = 0$, the one-point rule QRULE, LIKE9 with $m = 1$. The performance of LIKE9 improved steadily as m went from 1 to 9. For $m = 9$, its error rate was about one-half that of the one-point rule on the training sets, and on the test sets about three-fourths that of the one-point rule.

Qualitative comparison of the rules was made by using each to generate a map of a stretch of Imperial Valley data. Each rule was programmed with an option to allow a decision against all the alternative materials and to display such decisions by leaving blanks on the map. Such a "null test" creates a white framework of roads, rivers, and other extraneous materials against which materials of interest stand out. Large numbers of isolated recognitions made the one-point map difficult to read, but they were mostly removed by the nine-point rules. Incorrect classifications on the nine-point maps tended to occur in large patches. Fine detail on the one-point map, such as small roads, were lost or distorted on the nine-point maps.

LIKE9 and AVE9 have in common the premise that all nine pixels in the neighborhood represent the same material. Such rules would be expected to do well in field interiors where the premise holds and poorly on the boundaries where it doesn't. It is important that rules do well in boundary areas so that they will be effective on satellite data which contains a high proportion of boundary points. For this reason, the present study has been concerned with developing and testing nine-point rules based on assumptions less rigid than neighborhood conformity. In section 3, two such rules are derived and a third, arising naturally from the development of the second, defined. In section 4, these three rules are subjected to preliminary tests and compared in performance

with the previously-studied rules and the one-point rule. Section 5 reports the conclusions from these tests and recommends a plan for testing nine-point rules on LANDSAT data from the LACIE study. A glossary is provided at the end of the report to allow the reader to keep track of special names and symbols.

3

THREE NEW NINE-POINT RULES

3.1 BAYES9, A RULE BASED ON PARTIAL DEPENDENCE OF NEIGHBORING PIXELS

The classification rules previously studied have been based on one of two extreme assumptions:

1) The distribution of the data vector X is independent of the estimates of materials in neighboring pixels. This is the assumption underlying the usual classification rule.
2) The pixel and its neighbors represent the same material. This is the assumption on which the previously-studied nine-point rules are based.

The nine-point rules based on the second assumption did better than the one-point rule on field interiors but tended to clobber the boundary areas. The problem in the boundary areas seemed to be that the nine-point rules were sensitive to violations of the assumption on which they are based, as shown by their tendency to put a 3 x 3 white rectangle on a classification map whenever they encounter an odd data value. To correct such anomalies, a fix was programmed in the rules to omit the oddest point from the decision criterion. This takes care of the case where one point is non-conforming but what if, in a boundary area, several are?

A more fundamental reform is to return to the original intent of the nine-point rule and allow neighboring data values to influence but not dominate the center pixel decision. What is needed, in other words, is a rule based on an assumption flexible enough to tolerate non-conforming points in boundary areas - of special concern because of the goal of applying nine-point rules to satellite data in which boundary points are plentiful.

In this section, two rules BAYES9 and PRIOR9 based on flexible assumptions are derived. A third rule PREF9 based on the rigid assumption 2) is suggested by the derivation of PRIOR9.

BAYES9 is based on the assumption that a pixel <u>probably</u> represents the same material as neighboring pixels, a far more flexible assumption than the rigid requirement that it <u>certainly</u> conform. The flexible assumption, along with a technical assumption and Bayesian decision theory, are sufficient

to derive the decision rule BAYES9.* The derivation is given as Appendix A.

The rule depends on a constant θ that describes the degree of dependence between a pixel and its neighbor. The extreme values that θ can take are 0 for independence (assumption 1 above) and 1 for complete dependence (assumption 2).

To define θ more precisely, we let I and J be neighboring pixels. We assume prior probabilities P(I=a), where a is a typical material the pixel might represent and P stands for the operation "probability that". The conditional probability that I = a given that J = a, written symbolically P(I=a|J=a), is P(I=a) under an assumption of independence and is 1 under an assumption of complete dependence. The BAYES9 rule assumes an average of these two extreme values:

$$P(I=a|J=a) = (1-\theta)P(I=a) + \theta \cdot 1$$

When a and b are different materials, P(I=a|J=b) = P(I=a) under an assumption of independence but is 0 under an assumption of complete dependence. BAYES9 assumes the average of these two values:

$$P(I=a|J=b) = (1-\theta)P(I=a) + \theta \cdot 0$$

To summarize:

$$P(I=a|J=a) = (1-\theta)P(I=a) + \theta$$

$$P(I=a|J=b) = (1-\theta)P(I=a)$$

It is shown in Appendix A that this definition is consistent with the laws of probability.

When the prior probabilites are equal, the decision criterion is shown in Appendix A to be

$$P(X_0|a) \prod_{i=1}^{8} \left[P(X_i|a) + \frac{1-\theta}{\theta} \frac{\sum_b P(X_i|b)}{k}\right] \qquad (1)$$

*J. Gleason and M. McClary of ERIM originated the basic idea for this rule.

where

X_0 is the data value at the center pixel

$X_1, \ldots, X_8$ are the data values at the neighboring pixels

$P(X_i|a)$ is the probability density of X_i given material a

k is the number of materials

When $\theta = 1$, the second term drops out and the decision criterion is the product of the nine multivariate normal densities. This expression can be computed by summing the nine exponents and then exponentiating, so it is equivalent to the LIKE9 criterion. When $\theta \to 0$, the expression in the square brackets is asymptotic to the second term, which is the same for all materials, and thus the criterion reduces, essentially to $P(X_0|a)$, which is the criterion of the usual classification rule.

A suitable value of θ can be obtained from an empirical estimate p that neighboring pixels represent the same material, calculated by counting, on a recognition map, the number of pairs of neighboring pixels representing the same material and dividing by the total number of pairs.

Another estimate of p could be obtained from a geometrical probability calculation based on estimates of average field size and boundary width.

In the case of equal prior probabilities, θ would be obtained from p by solving

$$(1-\theta)\ 1/k + \theta = p$$

where θ is the BAYES9 parameter

k is the number of materials

p is the probability that two neighboring pixels represent the same material.

This relationship between θ, k and p is illustrated by Table 1.

TABLE 1. TABLE OF p FOR VARIOUS VALUES OF θ AND k.

	θ = .1	.3	.5	.7	.9
k = 2	.55	.65	.75	.85	.95
3	.40	.53	.67	.80	.93
4	.33	.48	.63	.78	.93
6	.25	.42	.58	.75	.92
8	.21	.39	.56	.74	.91
10	.19	.37	.55	.73	.91

The apparently cumbersome and time-consuming decision criterion (1) can actually be programmed to run quite rapidly. (The details are given in Appendix B). In a timing run on 10-channel data, 200 points per line, using a subset of six channels, the BAYES9 rule took only 11% longer than the usual maximum likelihood classification rule.

We run our classification rules with a null test (i.e., a test whether to decide "none of these") to allow for the existence in the scene of materials for which we don't have signatures. At first, with no clear rationale in mind, we used the test "whenever criterion (1) falls below a prescribed value, decide null". Maps made using this null test showed the same insensitivity to boundary detail and the same tendency to be splotched with 3 x 3 white rectangles as those made by LIKE9.

To correct this failing, a version BAYES8 of BAYES9 was programmed that left out the smallest square bracket factor in the BAYES9 decision criterion:

$$P(X_o|a) \prod_{i=1}^{8} [P(X_i|a) + \frac{1-\theta}{\theta} \sum_{b} \frac{P(X_i|b)}{k}] \qquad (1)$$

BAYES9 put out the number of the material with the smallest value of criterion (1) as channel 1 and the value itself, appropriately scaled, as channel 2 for use in the null test. A noisy pixel I_i would have small values $P(X_i|b)$ for all b. Thus the i^{th} square bracket term would be small and therefore criterion (1) would be small, signalling a null decision. The effect would occur whenever the pixel I_i were contained in the 3 x 3 neighborhood. Thus, a noisy pixel mapped by BAYES9 and its null test would produce a 3 x 3 white rectangle on the map.

When BAYES8 was defined to omit the smallest square bracket factor in criterion (1), an isolated noisy pixel triggered the null test only when it was the center pixel of the neighborhood and thus did not produce a white rectangle. Maps made by BAYES8 show the white rectangle program largely cured, but remain insensitive to fine detail on the boundaries, as evidenced by the omission and distortion of small boundaries observable on aerial photographs. Moreover, there was a slight loss of accuracy on field interiors

(see section 4). BAYES8 takes longer to run than BAYES9 - 16% longer than QRULE rather than 11% longer.

BAYES9 in theory should do better in the boundary areas than completely dependent nine-point rules. Because it did not, we looked to a poorly-defined null test as a likely culprit. Instead of blindly defining a null test by decision criterion (1) we derived one logically as follows.*

Whatever materials appear in the scene but are not given a specific distribution are lumped together in a null category, N, which is assumed to be distributed with a flat density of height ε. To simplify calculation and application, we assume that all materials and N have the same prior probability $1/(k+1)$. The BAYES9 criterion for each non-null material a is

$$P(X_o|a) \prod_{i=1}^{8} [P(X_i|a) + S \sum_b P(X_i|b)] \qquad (1)$$

where $S = \frac{1-\theta}{\theta(k+1)}$ and $\sum_b P(X_i|b)$ includes the null density ε. The criterion for the null material is

$$\varepsilon \prod_{i=1}^{8} [\varepsilon + S \sum_b P(X_i|b)] \qquad (2)$$

The decision procedure is to choose the material with the largest criterion (1) when this criterion exceeds the null criterion (2) and to choose null otherwise.

This definition of the null decision has two inconveniences. One is that the level ε of the null test must be set before the decision rule is run, whereas the null level of previous rules could be changed after the decision run, and so allow a map with too many or too few blank spaces to be rerun at a

*This rationale for defining a null test was suggested by R.J. Kauth of ERIM.

different level without redoing the decision run. This inconvenience can be remedied by defining two ε's. The first ε, used inside the square bracket factors, is set once and never changed. The ε outside the product sign, which we will call ε_2, can be adjusted later by putting the null test in the form "decide null if

$$\frac{\text{winning criterion (1)}}{\text{criterion (2)}} < 1"$$

i.e., if

$$\frac{\text{winning criterion (1)}}{\prod_{i=1}^{8} [\varepsilon + S \sum_{b} P(X_i|b)]} < \varepsilon_2$$

The left side, appropriately scaled, is stored on the processing output tape as channel 2 (the decision between materials is channel 1) and ε_2 is set as a control variable in the mapping or tallying module.

The other inconvenience is that it is not easy to estimate what value of ε to use. To solve a simpler problem first, let us find the density height ε that will reject one legitimate point in a thousand from a multivariate normal distribution with mean μ and covariance matrix R. The multivariate normal density is

$$c\, e^{-\frac{1}{2}[(X-\mu)^T R^{-1}(X-\mu) + \log_e|R|]}$$

Because c is the same for all materials, it will not be calculated for any of the material densities, and will therefore be omitted from further discussion. The quantity $Z = (X-\mu)^T R^{-1}(X-\mu)$ has a chi-square distribution. When Z is constant, the density is constant. Suppose we have four channels. From a table of the chi-square distribution we find that the probability that $Z > 18.465$ is .001. Let ε be the density when $Z = 18.465$. The probability of getting a data point X with a smaller density than ε is .001. Thus

$$\varepsilon = e^{-\frac{1}{2}(18.465 + \log_e|R|)}$$

Many signatures will produce many values $|R_i|$. We don't at the moment have clear advice about whether to choose the biggest $|R_i|$, the smallest $|R_i|$ or the average $|R_i|$ for the calculation of ε. Experience will have to be the guide.

The null test we have defined is to choose null when

$$\frac{P(X_o|a) \prod_{i=1}^{8} [P(X_i|a) + S \sum_b P(X_i|b)]}{\prod_{i=1}^{8} [\varepsilon + S \sum_b P(X_i|b)]} < \varepsilon_2$$

i.e., when

$$P(X_o|a) \prod_{i=1}^{8} \left[\frac{P(X_i|a) + S \sum_b P(X_i|b)}{\varepsilon + S \sum_b P(X_i|b)} \right] < \varepsilon_2 \qquad (3)$$

For the purpose of choosing among the non-null materials, criterion(3) is equivalent to criterion(1) because it is criterion(1) divided by an expression that is the same for all materials. The only additional computation required is one floating add per pixel and one floating divide per channel. And criterion(3) is the expression compared with a prescribed constant ε_2 in the null test. This test can be applied when making a map or when tallying the number of recognitions of each material.

BAYES9 with the better null test produced maps far more faithful to known boundaries than those of the completely dependent nine point rules. Moreover, fidelity increased, as it should, when θ decreased from .9 to .1 (see section 4.)

The decision rule is slightly changed by the improved null test because of the presence of ε in $\sum_b P(X_i|b)$. It should be a change for the better. When the data X_i from a neighboring pixel really represent material a, then $P(X_i|a)$ is large compared to ε and the presence of ε in $\sum_b P(X_i|b)$ makes virtually no difference. But if X_i is an odd data value, all $P(X_i|b)$'s are much smaller than ε and the square bracket factor is approximately $S/(1+S)$ for every material. Thus an odd point in the neighborhood would not make an unreliable contribution to criterion (3). In practice, results on field interiors (see section 4) are nearly identical for values of ε from 0 up to a value corresponding to a rejection level of .01.

3.2 PRIOR9, A BAYESIAN RULE BASED ON PRIOR PROBABILITIES DERIVED FROM NEIGHBORING DATA VALUES

The BAYES9 procedure of postulating a degree θ of dependence between neighboring pixels is one way to classify pixels with due regard for the influence but not domination of neighboring pixels. Another way is to use the data values of the 3 x 3 neighborhood to set prior probabilities for the decision on the center pixel. Such a procedure would embody the principle that the prior likelihood that a pixel represents a given material is dependent on the neighborhood in which the pixel is located. The classification rule PRIOR9 carries out this principle.*

Let the center pixel be I_o and the eight neighboring pixels $I_1,\ldots,I_8$. Let the corresponding data values be $X_o,\ldots,X_8$. We assume that the different materials a, b, c have global prior probabilities P(a), P(b), P(c). These can be set equal in the absence of other information. When the material, say a, is known, then the data values have known distributions, $P(X_o|a),\ldots,P(X_8|a)$, often assumed to be multivariate normal. By Bayes' formula, the posterior probability that a pixel I_i represents material a given the data value X_i is

$$P(a|X_i) = \frac{P(a)\ P(X_i|a)}{\sum_b P(b)\ P(X_i|b)}$$

Thus after looking at the data from the nine pixels, we have nine opinions $P(a|X_o),\ldots,P(a|X_8)$ of the probability that a pixel in the neighborhood is material a. They are consolidated into a single opinion by answering the question "What is the probability that if we pick a pixel at random from the neighborhood, it is material a?" This single number, which we take as the local prior probability of a when classifying the center pixel, is

$$\frac{1}{9} P(a|X_o) + \frac{1}{9} P(a|X_1) + \ldots + \frac{1}{9} P(a|X_8) \qquad (4)$$

*This decision rule was designed by J. Gleason and W. Richardson

Thus the decision criterion, writing out the Bayes' formulation of the posterior probabilities, is

$$P(X_o|a)\left[\frac{P(a)P(X_o|a)}{\sum P(b)P(X_o|b)} + \ldots + \frac{P(a)P(X_8|a)}{\sum P(b)P(X_8|b)}\right] \qquad (5)$$

(The factor, 1/9 is omitted because it is the same for all materials.)

This rule generalizes very simply to larger neighborhoods such as 5 x 5 and 7 x 7. If the whole scene were used as the neighborhood, we would have the one-point rule with two passes through the data to more accurately estimate the prior probabilities.

A purist might object to using the data X_o twice, once in the calculation of the local priors and again in the application of the decision rule. Two replies to this objection are:

1) Even if this rule were applied to a neighborhood of size one, the decision rule would be equivalent to the one-point rule because the material with the biggest weighted likelihood would get the biggest weight for the second application of the decision rule and would increase its lead over its rivals.
2) X_o must be represented in the calculation of the local prior probabilities to allow for the possibility that the center pixel is unlike its neighbors. If it were a small pond, for example, and the posterior probability of pond were essentially zero for the neighbors, the center pixel would have to participate in the prior probability to make possible a correct decision.

Because in practice many pixels do not represent a material searched for, the capability of deciding null (i.e., "none of these") is required for the accurate operation of the decision rule we have described. Otherwise, for such pixels meaningless posterior probabilities would be calculated and make unreliable contributions to the estimates of prior probabilities.

Our null test for PRIOR9 is based on the same principle as the one for BAYES9 (section 3.1). We define an additional category, the null category N, which has a flat distribution of height ε. N is chosen whenever criterion (5) for N

$$\varepsilon \left[\frac{P(N)}{P(N)\,\varepsilon + \sum_b P(b)P(X_o|b)} + \ldots + \frac{P(N)}{P(N)\varepsilon + \sum_b P(b)P(X_8|b)} \right] \quad (6)$$

is the largest. The prior probabilities are adjusted so that all of them including P(N) add up to 1.

The posterior probability of any material a

$$P(a|X_i) = \frac{P(a)\ P(X_i|a)}{\varepsilon P(N) + \sum_{\substack{\text{all} \\ \text{materials } b}} P(b)P(X_i|b)}$$

This posterior probability goes to zero for those pixels where all the material likelihoods are very small and so the unreliability of estimating the local prior probabilities is removed.

To simplify calculation and application, we assume that all materials and null have the same prior probability 1/(k+1) where k is the number of materials. Under these assumptions, the PRIOR9 criterion for each material a is

$$P(X_o|a) \left[\frac{P(X_o|a)}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{P(X_8|a)}{\varepsilon + \sum_b P(X_8|b)} \right] \quad (7)$$

and for the null category is

$$\varepsilon \left[\frac{\varepsilon}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{\varepsilon}{\varepsilon + \sum_b P(X_8|b)} \right] \quad (8)$$

The null test is to decide null if

$$\frac{\text{winning criterion (7)}}{\text{criterion (8)}} < 1$$

i.e., if

$$\frac{P(X_o|a)\left[\frac{P(X_o|a)}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{P(X_8|a)}{\varepsilon + \sum_b P(X_8|b)}\right]}{\frac{1}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{1}{\varepsilon + \sum_b P(X_8|b)}} < \varepsilon_2^2 \qquad (9)$$

The ε on the right side has been called ε_2 because it can be set at map-making time to regulate the amount of white space on the map, whereas ε must be set before applying PRIOR9 to the data values. The criterion can be thought of as the product of the center likelihood with a weighted average of the nine likelihoods. The weights, however, change from pixel to pixel.

PRIOR9 has been programmed as a processing module (Appendix C) and subjected to a preliminary test on aircraft data (section 4). Like BAYES9 it runs rapidly, taking only 12% longer than the usual one-point rule on a timing run on 10-channel data, 200 points per line, a subset of six channels used.

3.3 PREF9, AN IMPROVED VOTING RULE

The local prior probability of each material a

$$\frac{P(a)P(X_o|a)}{\sum_b P(b)P(X_o|b)} + \ldots + \frac{P(a)P(X_8|a)}{\sum_b P(b)P(X_8|b)} \qquad (10)$$

that was used by PRIOR9 in a Bayesian decision on the center pixel can be used alone as the criterion for a nine-point rule PREF9. The rule classifies the center pixel as the material giving the biggest value in answer to the question, "what is the probability that if you choose a pixel at random among the nine, it is material a?" The rule is similar in concept to the voting rule VOTE9

(a rule classifying the center pixel as the one with the most first place votes from the nine pixels) but bids fair to be more accurate because it uses all the information at each pixel and not just a vote for the winning material.

A null test for PREF9 is constructed by the same rationale as the one for BAYES9 and PRIOR9. We define a null category N having a flat distribution of height ε. N is chosen whenever criterion (10) for N

$$\frac{P(N)\varepsilon}{P(N)\varepsilon + \sum_b P(b)P(X_o|b)} + \ldots + \frac{P(N)\varepsilon}{P(N)\varepsilon + \sum_b P(b)P(X_8|b)} \qquad (11)$$

exceeds criterion (10) for every material. Let us assume all the materials and N have equal prior probabilities $1/(k+1)$. The test for null can be put in the form

$$\frac{\text{winning criterion (10)}}{\text{criterion (11)}} < 1$$

which, when written in detail, is

$$\frac{\dfrac{P(X_o|a)}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \dfrac{P(X_8|a)}{\varepsilon + \sum_b P(X_8|b)}}{\dfrac{\varepsilon}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \dfrac{\varepsilon}{\varepsilon + \sum_b P(X_8|b)}} < 1$$

in other words

$$\frac{\dfrac{P(X_o|a)}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \dfrac{P(X_8|a)}{\varepsilon + \sum_b P(X_8|b)}}{\dfrac{1}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \dfrac{1}{\varepsilon + \sum_b P(X_8|b)}} < \varepsilon_2$$

Theoretically, ϵ_2 is the prescribed null test level ϵ. But because ϵ_2 can be changed at map-making time, the practical course is to set ϵ as best we can, run the decision rule PREF9, and then adjust ϵ_2 to produce a suitable amount of white space on the map.

Because the BAYES9 and PRIOR9 decisions are primarily controlled by the center pixel values, we would expect these rules to be more sensitive to fine detail than PREF9, for which the criterion is a function of all nine points of the neighborhood without particular regard for the center pixel. For this reason, of all the nine-point rules so far defined, PRIOR9 and BAYES9 have the best hope of attaining the high resolution critical for processing LANDSAT data. We would expect all three rules to be relatively insensitive to the presence of noisy pixels in the neighborhood because very small likelihoods are dominated by ϵ in the decision criteria.

The calculations to carry out PREF9 are nearly all required for PRIOR9 (see Appendix C). For this reason both rules are implemented by the same processing module and an imput switch determines whether one rule or the other or both will be used. The timing run for PRIOR9, mentioned earlier, that took 12% longer than the one-point rule was actually a run of PRIOR9 and PREF9 together. A neglible amount of time would have been saved by running one of them alone.

4

PRELIMINARY TESTS OF NINE-POINT RULES

In this section, we report a preliminary test of the new rules BAYES9, PRIOR9 and PREF9 and a test of BAYES9 and the previously-studied rules LIKE9, AVE9 and VOTE9[1] on a new data set.

The three nine-point decision rules introduced in this report, BAYES9, PRIOR9 and PREF9, were given preliminary tests on aircraft data collected from the Imperial Valley, California. This data set was used because 1) we have confidence in the ground truth [2, pp. 149-172], 2) the crops are not easy to distinguish, so that differences in performance of the rules may be displayed, 3) we have identified 42 fields for which the ground truth is unequivocal and the scan angle minimal, a number of replications sufficient for experimental comparison, 4) the previously-studied rules were tested on this data set and the results are available for comparison, 5) the field boundaries are easily-recognizable lines verified by aerial photographs, so that we can observe on maps made by the rules how well the rules are doing on the boundaries. The next stage of testing will be on the LACIE intensive study site data which was just becoming available when this report was written.

The first test was on field interiors from the 42 fields. Signatures for alfalfa, barley, lettuce, rye, bare soil, sugar beets and saffalowers were computed from a set SIGS 1 of 20 training fields and then from another, disjoint set of 20 training fields SIGS 2. Two runs were made using SIGS 1 and SIGS2, respectively. The per cent misclassified by the rules tested is reported in Table 2. Results for LIKE9 with m=9, AVE9 with t=1 and VOTE9 are included for a comparison.

---

[1] W. Richardson, A Study of Some Nine-Element Decision Rules, Technical Report 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 1974.

[2] R.F. Nalepka, Investigation of Multispectral Discrimination Techniques Technical Report 2264,12-F, Willow Run Laboratories, Ann Arbor, Michigan, 1970.

The "level" associated with BAYES9, PRIOR9 and PREF9 is the percentage of a typical material distribution where the density is less than the flat density of height ε in the null test (see section 3.1). To run one of these rules with a level of .001, say, you look up the number in a chi-square table at the .001 column and row 6 (if there were six original data channels), obtaining the number 22.457. You get the average $\log_e |R_i|$, say 2.57, where R_i is the covariance matrix of material i, and include in the input the statements

EXPLIM = 22.457, LOGR = 2.57

$\text{Log}_e |R_i|$ is included in the output of the preceding module QRULE. The nine-point module computes

$$\varepsilon = e^{-\frac{1}{2}(\text{EXPLIM} + \text{LOGR})}$$

A level of 0 corresponds to an ε of 0. The BAYES9 results for ε=0 were obtained by an early version of the module that is logically equivalent to the current version when ε=0 and the null test isn't used (as it wasn't in the field center experiment).

BAYES9 and the three previously-studied rules LIKE9, AVE9, and VOTE9 were tested on field interiors from four segments of the Corn Blight Watch Experiment in Indiana, 1971. A total of 225 fields were included. Seven materials were recognized: corn, soybeans, trees, pasture, hay, growing hay, and bare soil. For each material, every other field was selected as a training field in the order the fields appeared on the tape. This sampling scheme provided a class of training fields large enough to achieve meaningful training field tests of the decision rules. The choice of every other field avoided the possible inference that a peculiar grouping of training fields affected the results.

A summary of the results is given in Tables 3-5. Imperial Valley results reported in Table 1 and [1] are included for comparison. The error rate reported is an average of the error rates of the individual fields.

---

[1] W. Richardson, A Study of Some Nine-Element Decision Rules, Technical Report 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 1974.

TABLE 2. PERCENT MISCLASSIFIED BY VARIOUS DECISION RULES ON 42 FIELDS FROM THE IMPERIAL VALLEY USING TWO DIFFERENT SELECTIONS, SIGS 1 AND SIGS 2, OF TRAINING FIELDS

Decision Rule	20 Training Fields SIGS 1	20 Training Fields SIGS 2	22 Test Fields SIGS 1	22 Test Fields SIGS 2
QRULE (usual rule)	20.3	21.8	31.6	32.8
PRIOR9, level = .001	16.4	17.1	28.9	29.0
PREF9, level = .001	11.9	12.2	25.2	24.6
VOTE9	13.5	14.2	26.8	26.7
BAYES9, level = 0, θ = .1	13.7	14.0	27.2	25.7
θ = .3	12.3	12.2	25.6	24.4
θ = .5	11.9	11.7	24.8	23.9
θ = .7	11.7	11.4	24.5	23.6
θ = .9	11.6	11.1	24.3	23.4
BAYES9, θ = .5, level = 0*	11.85	11.70	24.81	23.90
level = .001	11.87	11.61	24.75	23.93
level = .01	11.81	11.62	24.76	24.07
level = .1	11.98	12.16	24.87	24.69
BAYES8, θ = .5	12.6	13.0	25.4	25.0
LIKE9, m = 9	11.7	11.4	24.8	23.5
AVE9, t = 1	14.1	17.6	26.2	28.9

*This line is repeated for ready comparison with the other levels.

TABLE 3. PERCENT MISCLASSIFIED BY FIVE TYPES OF DECISION RULES ON TRAINING FIELDS FROM THE IMPERIAL VALLEY (TWO SETS OF SIGNATURES) AND FROM FOUR SEGMENTS OF THE CORN BLIGHT WATCH EXPERIMENT.

	Imperial Valley		Corn Blight Watch			
	SIGS 1	SIGS 2	S203	S204	S212	S227
Number of Fields:	20	20	17	34	27	30
QRULE (Usual Rule)	20.3	21.8	5.6	4.7	7.0	5.8
LIKE9, m = 1	20.0	26.1	4.2	3.1	6.9	4.9
m = 3	16.1	21.9	3.2	2.7	6.1	4.8
m = 5	14.1	17.6	2.8	2.7	5.3	5.2
m = 7	12.7	13.3	2.8	2.7	5.0	4.3
m = 8	12.0	12.3	3.1	2.8	4.7	4.5
m = 9	11.7	11.4	3.2	3.0	4.7	4.4
AVE9, t = 0	15.5	18.9	3.3	2.9	5.5	4.4
t = 1	14.1	17.6	3.3	2.9	5.5	3.9
t = 3	13.7	17.6	3.0	3.0	5.3	4.4
VOTE9	13.5	14.2	2.7	3.1	5.2	4.2
BAYE9, level = 0						
θ = .1	13.7	14.0	3.5	3.5	5.5	4.9
θ = .3	12.3	12.2	3.1	3.2	5.0	4.1
θ = .5	11.9	11.7	3.0	3.2	4.9	4.0
θ = .7	11.7	11.4	2.9	3.1	4.8	4.0
θ = .9	11.6	11.1	2.9	3.0	4.8	3.7
BAYES8, θ = .5	12.6	13.0	3.1	3.1	5.1	4.5

TABLE 4. PERCENT MISCLASSIFIED BY FIVE TYPES OF DECISION RULES ON TEST FIELDS FROM THE IMPERIAL VALLEY (TWO SETS OF SIGNATURES) AND FROM FOUR SEGMENTS OF THE CORN BLIGHT WATCH EXPERIMENT.

	Imperial Valley		Corn Blight Watch			
	SIGS 1	SIGS 2	S203	S204	S212	S227
Number of Fields:	22	22	17	36	30	34
QRULE (Usual Rule)	31.6	32.8	10.5	10.1	11.5	15.7
LIKE9, m = 1	34.1	35.7	10.5	7.0	9.0	14.9
m = 3	30.0	32.1	9.1	7.3	8.4	13.9
m = 5	27.3	28.8	8.8	7.9	8.4	13.5
m = 7	25.5	25.6	8.7	8.7	8.9	13.3
m = 8	25.0	24.1	8.8	9.1	9.1	13.4
m = 9	24.8	23.5	9.1	9.6	9.3	12.8
AVE9, t = 0	27.5	29.9	9.9	9.5	9.6	13.2
t = 1	26.2	28.9	9.8	9.4	9.5	13.5
t = 3	26.2	28.4	9.6	9.3	9.3	13.7
VOTE9	26.8	26.7	8.4	8.7	9.1	13.8
BAYE9, level = 0						
θ = .1	27.2	25.7	9.3	9.2	9.4	14.1
θ = .3	25.6	24.4	8.7	9.0	8.9	13.2
θ = .5	24.8	23.9	8.6	8.9	8.8	13.1
θ = .7	24.5	23.6	8.5	8.9	8.7	13.0
θ = .9	24.3	23.4	8.4	8.9	8.6	12.8
BAYES8, θ = .5	25.4	25.0	8.7	8.8	8.7	13.4

TABLE 5. ERROR RATES FOR FIVE TYPES OF DECISION RULES AVERAGED OVER FOUR SEGMENTS OF THE CORN BLIGHT WATCH EXPERIMENT

	Training Fields	Test Fields
QRULE (Usual Rule)	5.77	11.96
LIKE9, m = 1	4.78	10.32
m = 3	4.19	9.66
m = 5	3.98	9.67
m = 7	3.72	9.89
m = 8	3.80	10.08
m = 9	3.81	10.20
AVE9, t = 0	4.03	10.53
t = 1	3.88	10.55
t = 3	3.94	10.47
VOTE9	3.81	10.00
BAYES 9, level = 0		
θ = .1	4.35	10.48
θ = .3	3.86	9.94
θ = .5	3.76	9.84
θ = .7	3.71	9.76
θ = .9	3.61	9.69
BAYES8, θ = .5	3.93	9.91

TABLE 6. PERCENT MISCLASSIFIED BY THE USUAL DECISION RULE ON TEST FIELDS FOR HAY, GROWING HAY AND PASTURE. THE COUNT IN THE FIRST COLUMN INCLUDES EVERY MISIDENTIFICATION; IN THE SECOND COLUMN, WRONG CHOICES AMONG THE THREE MATERIALS ARE NOT COUNTED

	INDIVIDUAL	3 MATERIALS GROUPED
PAST 25	0	0
PAST 27	78	76
PAST 86	22	7
PAST 88	52	33
HAY 90	78	28
HAY 92	81	81
GHAY 94	24	22
GHAY 96	16	11
GHAY 98	54	38
GHAY 100	3	0
PAST 139	20	0
PAST 141	35	14
PAST 143	65	2
PAST 145	99	0
PAST 147	4	4
HAY 148	96	75
HAY 150	99	12
GHAY 152	84	16
GHAY 154	0	0
PAST 201	99	96
PAST 203	9	0
PAST 205	78	1
PAST 207	26	23
HAY 208	68	0
HAY 210	55	55
HAY 214	96	77
GHAY 215	99	1
GHAY 217	3	0

TABLE 7. NUMBER OF DECISIONS MADE JOINTLY BY QRULE, PRIOR9 AND PREF9 ON 42 IMPERIAL VALLEY FIELDS USING THE FIRST SET (SIGS 1) OF TRAINING FIELDS

	20 Training Fields		22 Test Fields	
	Correct	Incorrect	Correct	Incorrect
QRULE alone	149	922	218	811
PRIOR9 alone	26	209	34	97
PREF9 alone	834	491	872	737
QRULE with PRIOR9	241	857	427	1062
QRULE with PREF9	7	1	9	2
PRIOR9 with PREF9	587	257	545	364
all three together	9278	1173	9169	2072

The signatures for pasture, hay, and growing hay extended very poorly from training fields to test fields, as is shown in Table 6. To avoid letting these huge errors dominate the experiment, and mindful that the three materials were quite similar, we did not count an error if one of these three materials were recognized as another. If one of these materials were recognized as corn, say, then that was counted as an error, because we were seeing corn where there wasn't any. And, of course, mistaking a corn pixel for one of the three materials was also an error.

The test on the Corn Blight data (Tables 3 and 4) confirmed the previous finding, reported in columns 1 and 2, that the three previously-studied nine-point rules have lower error rates than the one-point rule and that the per cent improvement is greater for the training than for the test sets. These results were obtained on the interiors of fields where the assumption underlying the three rules, namely that a pixel represents the same material as its immediate neighbors, would be expected to hold. Yet surprisingly, the BAYES9 results for $\theta = .3$ through $\theta = .9$, were just as good as the other results. In the Imperial Valley tests the BAYES9 results for $\theta = .3$ through $\theta = .7$ are about the same as for LIKE9 with $m = 8$ and 9, and better than all the other results. The Corn Blight results were more contradictory, but the BAYES9 rule always did well, and if one looks at the results averaged over all four segments (Table 5) one finds the BAYES9 results, except for $\theta = .1$, are best for the training fields and a close second to LIKE9 with $m = 3$ and $m = 5$ on the test fields. A possible explanation for the success of BAYES9 on field interiors is that even when one expects uniformity, there may be exceptional pixels or noisily recorded data where the probabilitic assumption is more applicable than the extreme assumption

In all six of the runs, the error rate for BAYES9 decreased substantially when θ went from .1 to .3, and although there was no substantial difference in the results for $\theta = .3 - .9$, the rates decreased very gradually as θ increased.

The Corn Blight results do not present the same consistent picture as the Imperial Valley results. The latter results showed a steady improvement as m increased from 1 to 9. In the Corn Blight segments 203 and 204, the best-m-of-nine rule LIKE9 did best for middle values of m. In the test field results of segment 204, the error rate worsened as m went from 1 to 9. The voting rule VOTE9 was best in segment 203 and did fairly well in the other runs but was not exceptional. The trimmed mean rule AVE9 generally didn't do as well as the others, but did well on the training fields of segment 204 and 227. LIKE9 with m = 9 was a little worse than m = 8 for all segments except 227. The BAYES9 rule for $\theta = .7$ and $\theta = .9$ was best on segment 227 and the test fields of 203.

A look at the error rates averaged over all four Corn Blight segments (Table 5) shows that the BAYES9 rule with $\theta = .9$ made 2/3 the errors of the one-point rule on the training fields and 4/5 on the test fields. For the training fields, BAYES9 with $\theta = .5$ through .9, LIKE9 with m = 7 through 9 and VOTE9 had the lowest overall rates. For the test fields, BAYES 9 with $\theta = .5$ through .9 and LIKE9 with m = 3 through 7 had the lowest rates. BAYES8 generally did a little worse than BAYES9 with a comparable setting of θ.

Of the two voting rules, PREF9 did consistently a little better than VOTE9 on field interiors from the Imperial Valley data (Table 2) and is comparable in performance to the best of the other nine-point rules.

The PRIOR9 test results fell about half way between the PREF9 and QRULE results. This is not surprising when you consider that the PRIOR9 decision criterion $p_{oa} \sum_{i=o}^{8} p_{ia}$ is equivalent to the geometric mean between the PREF9 criterion $\sum_{i=o}^{8} p_{ia}$ and an expression p_{oa} equivalent to the QRULE decision criterion (see Appendix C for a definition of these terms).

A small module was run to count the number of agreements and disagreements between QRULE, PRIOR9, PREF9 and the ground truth. The results are given as Table 7. The table shows that when QRULE and PREF9 disagree, PRIOR9 is slightly more likely to side with QRULE than PREF9, and when it

does it is far more likely to be wrong. When all three rules agree, they are very likely to be right. The numbers in the table don't add up to the total number of pixels because when PRIOR9 sides with QRULE, for example, then PREF9 is alone and so that category is also incremented. All of the categories containing PREF9 add to the total number of pixels and that statement holds also for the other two rules. Theoretically, there should be no entries in the QRULE plus PREF9 category; a different method of handling ties must account for non-zero numbers there.

PRIOR9 performed as well as the moving average rule AVE9 on the training and test sets of the SIGS 2 run, and only slightly worse on the SIGS 1 run.

The BAYES decision criterion

$$P(X_o|a) \prod_{i=1}^{8} \left[\frac{P(X_i|a) + S\varepsilon + \sum_b P(X_i|b)}{\varepsilon + S\varepsilon + \sum_b P(X_i|b)} \right]$$

gives slightly differing results for differing values of ε. When $\varepsilon > 0$, small probabilities $P(X_i|a)$ are overshadowed by ε and so the square bracket factor corresponding to an oddball point tends to make the same contribution for all materials. This should be preferable to the tendency of the $\varepsilon = 0$ rule to let every oddball point in the neighborhood exert an erratic influence over the decision. Conversely, too large a value of ε might tend to impose an unnecessary uniformity on square bracket factors that would otherwise help distinguish between materials in a valid way. The test results in Table 2 however, show almost no difference in results for the the levels 0, .001, .01 and even .1. In both test and training fields, the level .1 is very slightly worse.

A second test of the new rules was to use them to make maps of a stretch of Imperial Valley data. Portions of these maps are given as Figures 2 through 6 in the order PRIOR9, PREF9, BAYES9 with $\theta = .1$, with $\theta = .3$ and with $\theta = .9$. (The level in every case was .001.) For comparison a QRULE map is given as Figure 1 and BAYES8 map as Figure 7.

The purpose of the map making was to compare the performance of the rules on the boundary areas. The comparison is necessary because rules such as those assuming conformity of the nine pixels might do well on field interiors where the assumption holds and do badly on the boundaries where the assumption fails. The comparison is qualitative rather than quantitative because we can't be sure of the ground truth of any boundary pixel. We do know from aerial photographs [2, p. 155] the pattern of field boundaries and observe that they are in many cases traced out or suggested in the QRULE (usual one-point rule) map. Rules that clobber the boundaries as observed on the QRULE map or omit many sections of boundaries are, we suspect, classifying poorly in the boundary areas. Rules that preserve fine detail on the boundaries are likely to be classifying more accurately there.

To make the maps comparable, a processing module HISTBN was written that prints the cumulative distribution of a selected channel as a table of percentages. When we look at such a table of the null test criterion, we are able to determine a null test level that will produce a desired percentage of white space on the map. From the statistics of previous maps of the area, we determined tha 11% white space produced clearly delineated boundaries while leaving intact the field interiors. By using HISTBN, we were able to print the seven maps with 11% white space.

The map of PREF9 (Figure 3) although it appears to be the neatest because of the smoothing tendency of the PREF9 rule, is the least faithful to fine detail on the boundary. Many sections of boundary observable on the QRULE map (Figure 1) are missing. The boundaries that do appear are distorted by white blobs.

The PRIOR9 map (Figure 2) is very faithful to the fine detail on the boundaries. Hardly any sections of boundary suggested on the QRULE map are missing. The boundaries are the same reasonable shape as on the

---

[2] R.F. Nalepka, Investigation of Multispectral Discrimination Techniques Technical Report 2264,12-F, Willow Run Laboratories, Ann Arbor, Michigan, 1970.

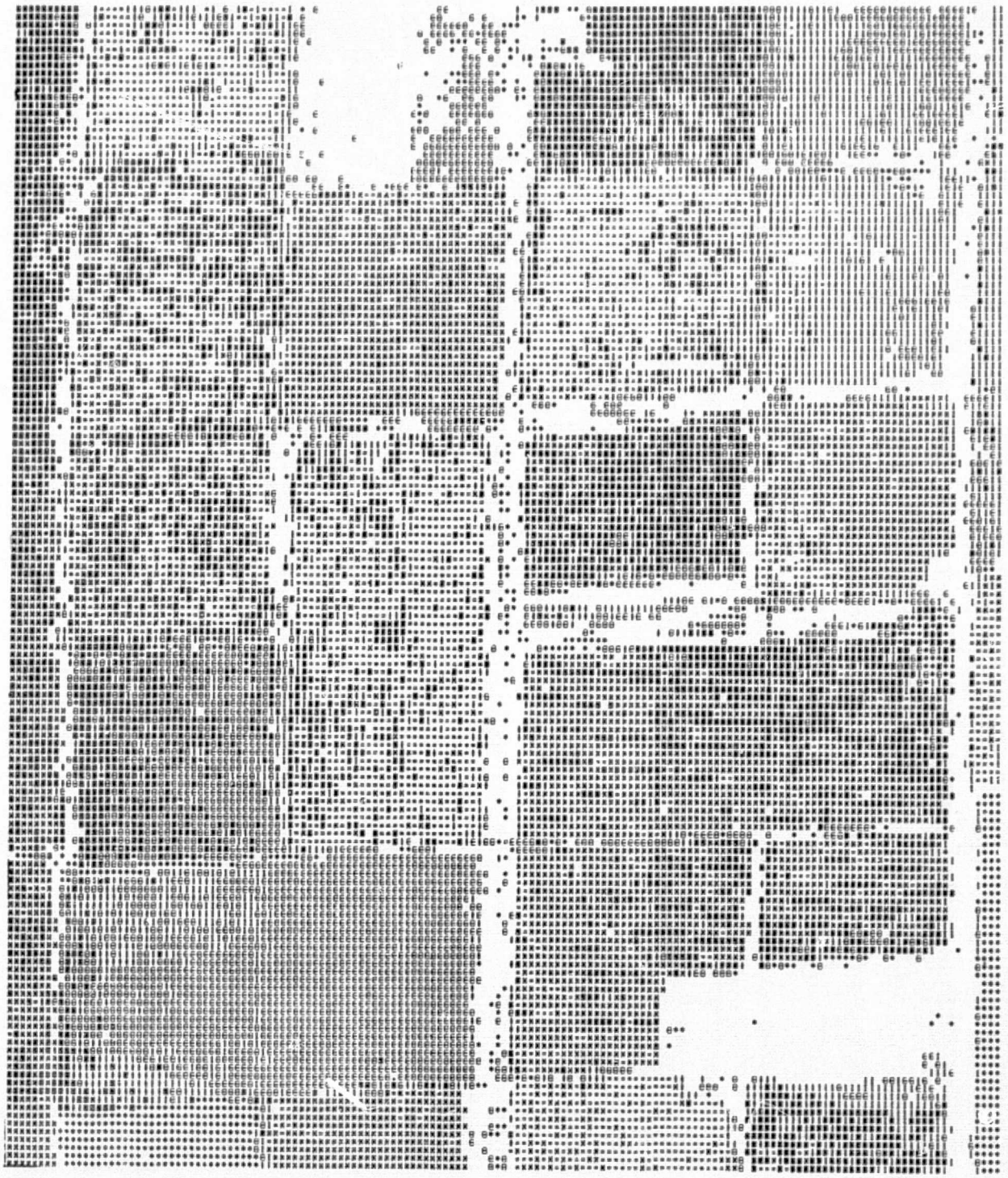


FIGURE 1. QRULE, THE USUAL ONE-POINT RULE

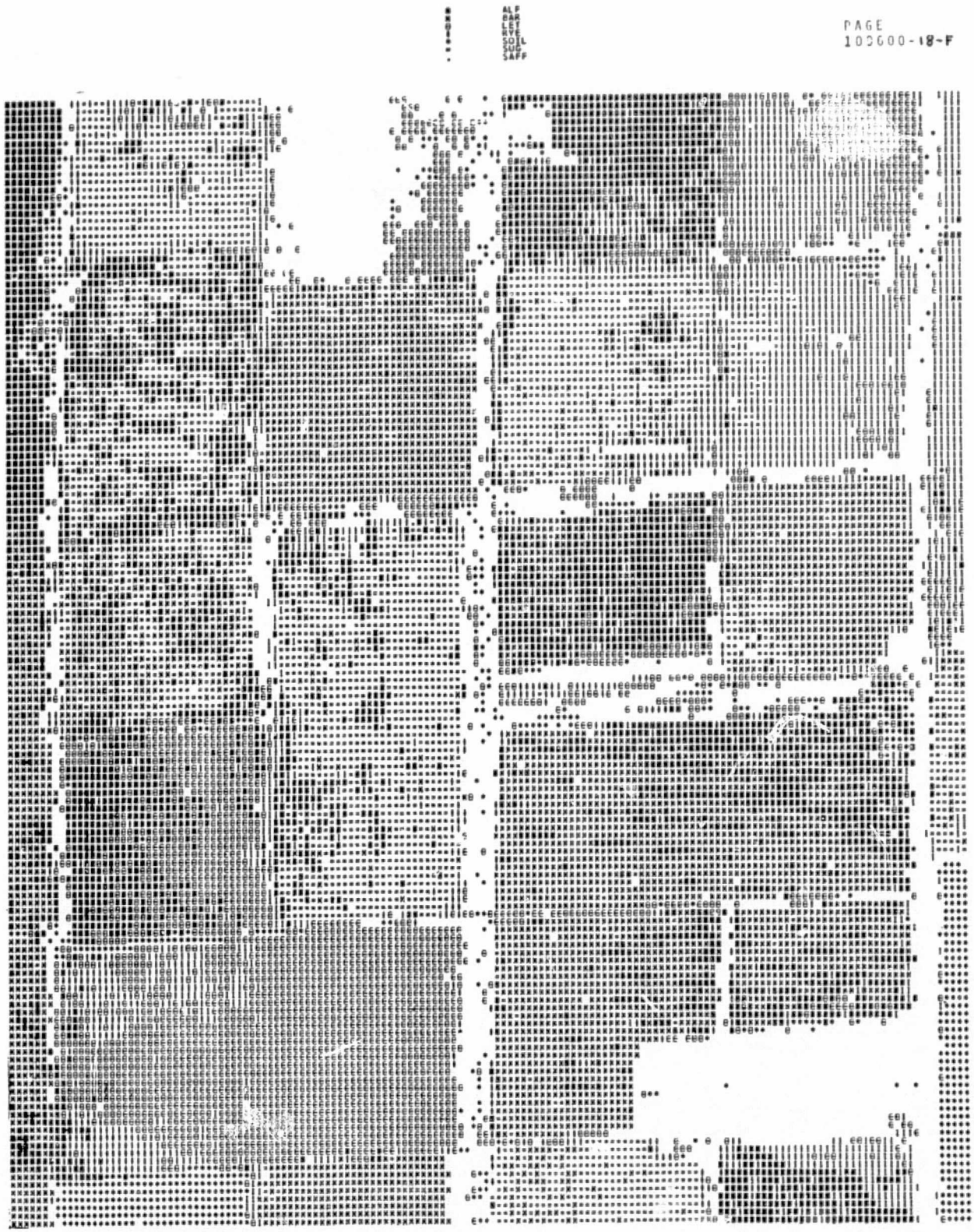


FIGURE 2. PRIOR9, A BAYESIAN RULE WHOSE PRIOR PROBABILITIES ARE DERIVED FROM NEIGHBORING DATA VALUES

FIGURE 3. PREF9, AN IMPROVED VOTING RULE BASED ON POSTERIOR PROBABILITIES SUMMED OVER THE NEIGHBORHOOD

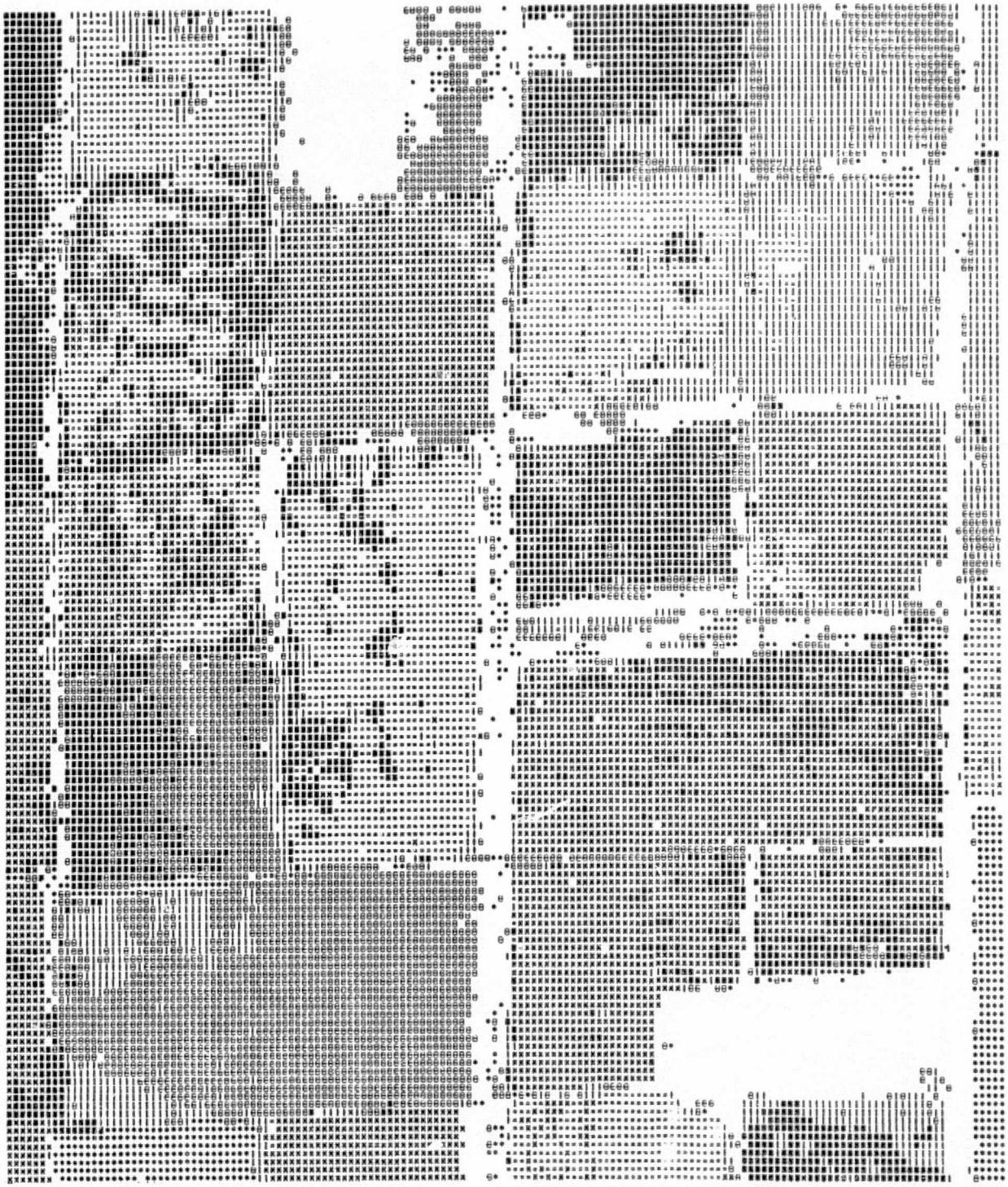


FIGURE 4. BAYES9, θ= .1, A RULE BASED ON PARTIAL DEPENDENCE OF NEIGHBORING PIXELS, DEPENDENCE PARAMETER = .1.

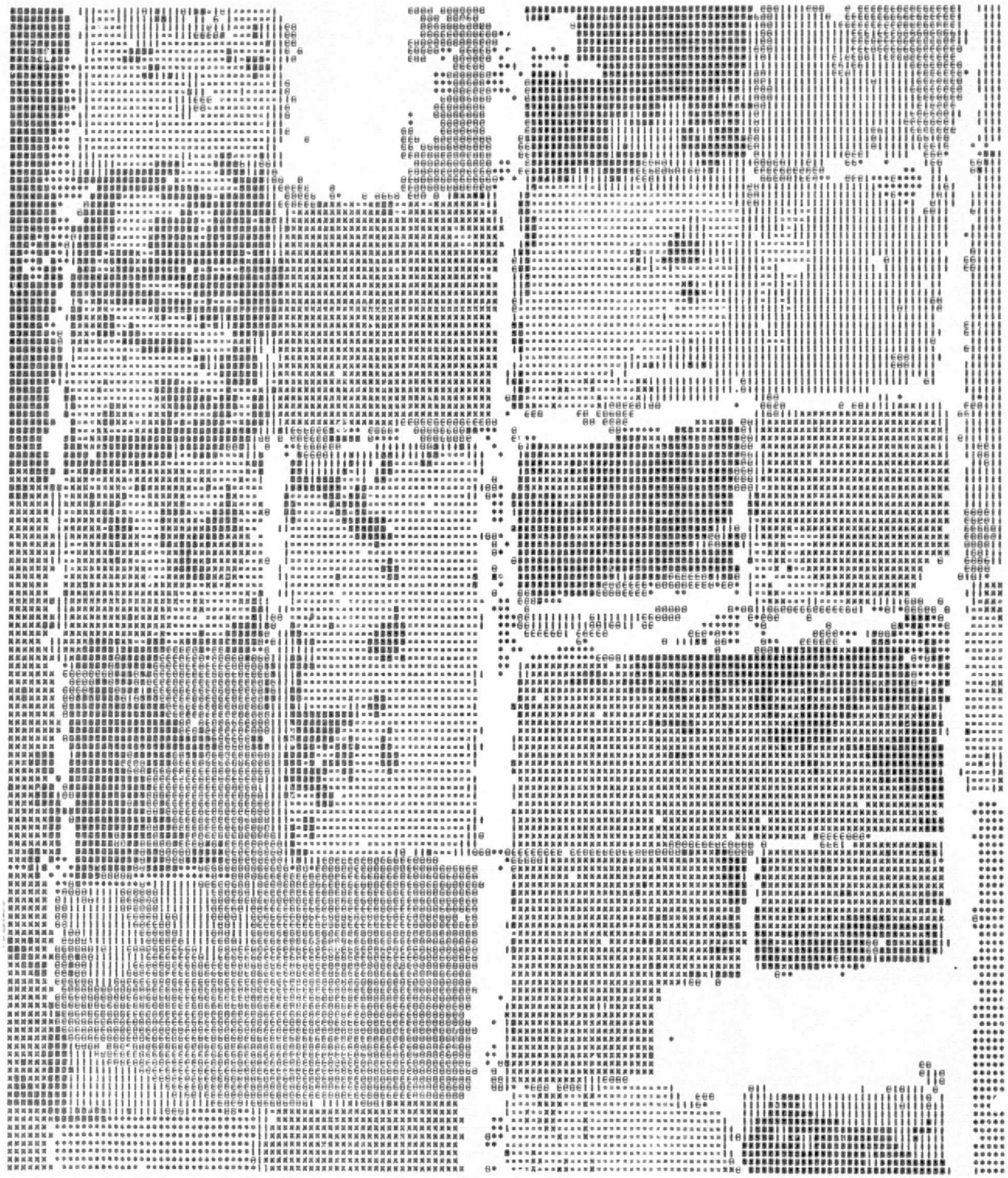


FIGURE 5. BAYES9, θ = .3, A RULE BASED ON PARTIAL DEPENDENCE OF NEIGHBORING PIXELS, DEPENDENCE PARAMETER = .3.

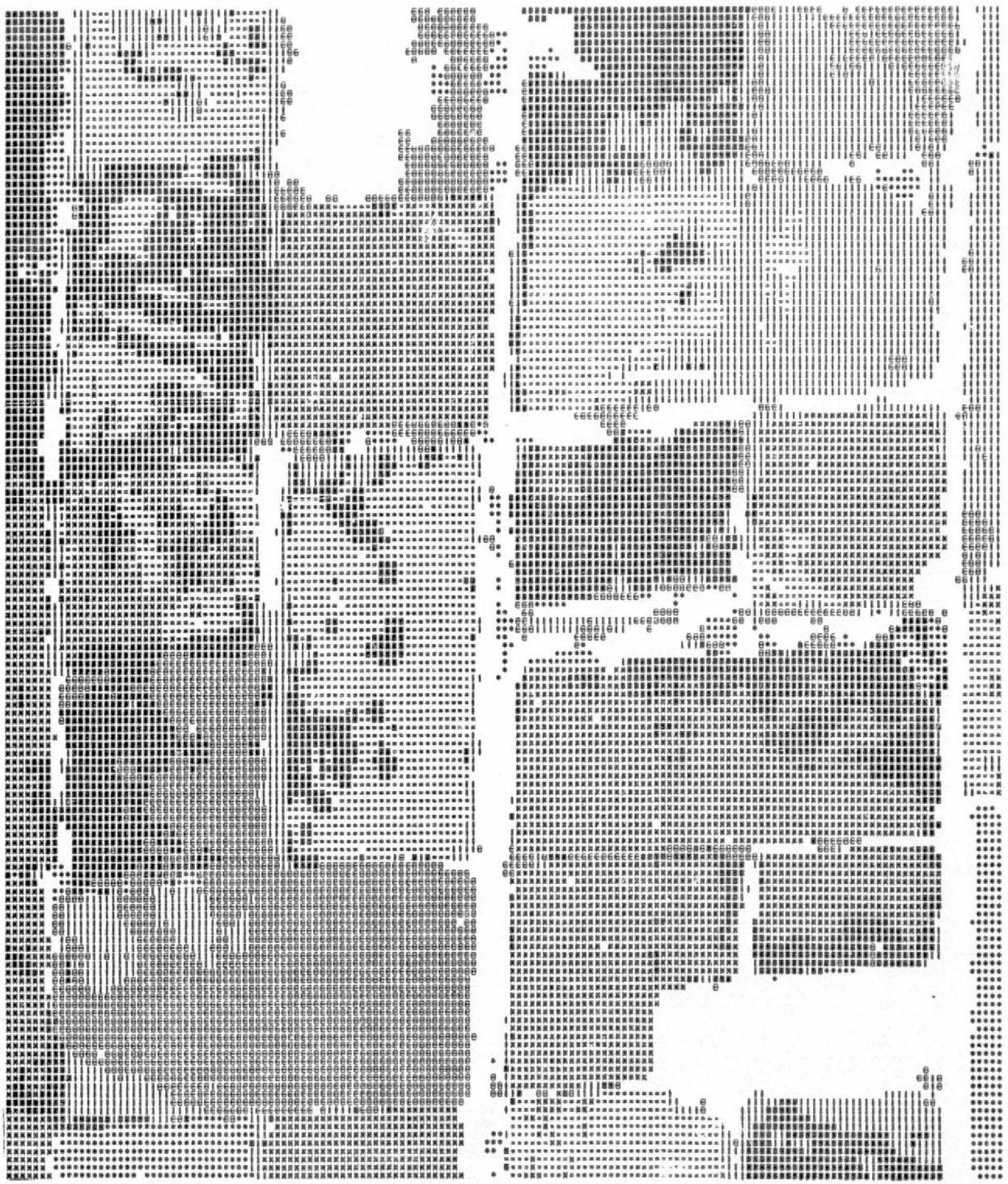


FIGURE 6. BAYES9, θ = .9, A RULE BASED ON PARTIAL DEPENDENCE OF NEIGHBORING PIXELS, DEPENDENCE PARAMETER = .9.

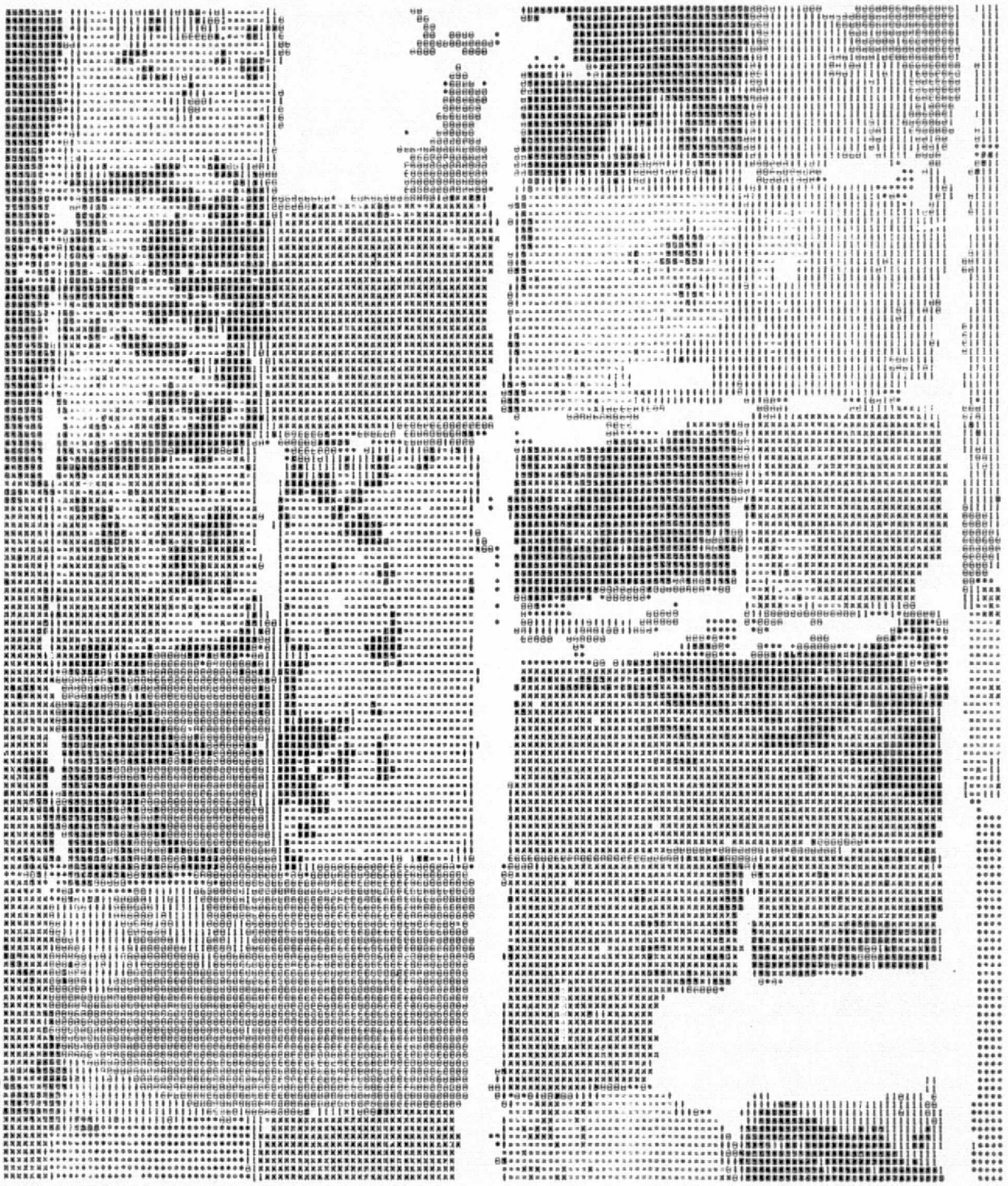


FIGURE 7. BAYES8, θ = .3, A RULE LIKE BAYES9 BUT WITH THE ODDEST NEIGHBOR OMITTED.

QRULE map; they are not distorted. The PRIOR9 map is also a good bit more speckled than the PREF9 map, reflecting its poorer performance on field interiors.

The BAYES9 maps (Figures 4-6) show a gradation of performance. As θ goes from .1 to .3 to .9, small boundaries are lost, the boundaries get more distorted and the map less speckled. The map for $\theta = .3$ appears to be a good compromise. The distortion is small, the speckling is slight and the small boundaries are mostly there. The $\theta = .9$ map is more faithful to the boundaries than the PREF9 map and slightly more speckled.

The BAYES8 map (Figure 7), although an improvement over BAYES9 maps made with the old null test, is considerably less faithful to boundary detail than the new BAYES9 maps (Figures 4-6), showing that the new null test rather than the BAYES8 approach was what was needed to demonstrate performance in the boundary areas.

The maps of LIKE9 and AVE9 given in [1, pp. 28-31] are similar in appearance to the PREF9 and BAYES8 maps with respect to their neat appearance and their omission and distortion of boundaries.

The map of VOTE9 [1, p.32] picked up many small boundaries missed by the other rules and some, even, missed by QRULE. This is because the null test criterion (as well as the decision criterion) is the number of first place votes for a material. When this fell below a prescribed number, a blank was printed. Thus when the 3 x 3 neighborhood fell on top of a boundary between two materials, even if there were no noticeable boundary strip such as a road between them, the disputed vote resulted in a null decision. However, many extraneous areas not identifiable with one of the materials were filled in rather than being left blank as with the other rules. Also, many interior points were left blank. Thus the VOTE9 null test was more of a boundary detector than a null test and is not comparable to the null tests of the other sides.

---

[1] W. Richardson, A Study of Some Nine-Element Decision Rules, Technical Report 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 1974.

5
CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

The following conclusions are tentative because they are drawn from tests on two aircraft data sets. Further testing on satellite data will be required to establish their general validity.

BAYES9 is a nine-point rule that combines accuracy in field interiors with sensitivity to detail on the boundaries. Although based on an assumption of only partial dependence, it performed as well on field interiors as any of the rules based on the presumably valid assumption of complete dependence and substantially exceeded the performance of the one-point rule, QRULE. In boundary areas, BAYES9 showed much less distortion and omission of boundary detail than the rules based on complete dependence. The increase in fidelity to detail as the BAYES9 dependence parameter θ went from .9 to .1 shows that BAYES9 can be adjusted to get the best tradeoff between boundary and field center performance. It is therefore a promising rule to use on LANDSAT data which has a high proportion of boundary points and for which the best balance between field interior and boundary performance has yet to be determined.

PRIOR9, a Bayesian rule based on prior probabilities derived from neighboring data values, is particularly effective in boundary areas and outperforms QRULE on field interiors, although the margin of improvement is about half that of BAYES9, LIKE9, and PREF9. Because LANDSAT data contains a high proportion of boundary points, PRIOR9 merits testing on LANDSAT data.

PREF9 is a rule based on posterior probabilities summed over the 3 x 3 neighborhood. It is therefore based on the assumption of complete dependence (i.e., the assumption that all nine pixels represent the same material) and exhibits the strengths and weaknesses of such rules: good performance on field interiors and insensitive performance on the boundaries. PREF9 resembles the voting rule VOTE9 except that it uses all the information at each pixel of the neighborhood rather than just a vote for the winning material. It is thus theoretically more appealing than VOTE9 and it performed better in the

the preliminary tests. The margin of improvement was slight, and so the conclusion of superior performance is tentative. Because VOTE9 is analogous to counting recognitions to estimate acreage and PREF9 is analogous to the method of summing posterior probabilities, we are led to suspect that the latter method of acreage estimation is more accurate.

The success of the performance of BAYES9 and PRIOR9 in boundary areas speaks well for the null test concept that they both use, namely, that materials appearing in the scene but not given a specific distribution can be lumped together in a null category which has a flat distribution of height ε. Theoretically, such a null test should perform best if ε is greater than zero but not too large. Judging from our preliminary test of BAYES9 at four rejection levels, the effect is very slight. Only at a rejection level of 10% was there even a slight upturn in rates.

BAYES8, intended to fix BAYES9 to improve performance on the boundaries, is less faithful to fine detail on the boundaries than BAYES9 with the new null test, has poorer performance on field interiors, takes longer and does not have the sound theoretical justification of BAYES9. We conclude that BAYES8 is not sufficiently promising for further testing.

The moving average rule AVE9 compares unfavorably with PRIOR9 - hardly any better on field interiors and far worse on the boundaries. The chief value of AVE9 and VOTE9 is ease of calculation. These two modules could be run with the best linear rule [3] and thus save considerable time. If they were run with the usual one-point rule QRULE, they would not be much faster than BAYES9 or PRIOR9 which took 11% and 12% longer, respectively, than QRULE alone in a six-channel timing run.

5.2 RECOMMENDATIONS

The development of nine-point classification rules that substantially reduced the error rate on field interiors from aircraft data and also appeared to classify accurately in boundary areas encourages the hope that these rules will generally

[3] R.B. Crane and W. Richardson, Performance Evaluation of Multispectral Scanner Classification Techniques, Proceedings of the 8th International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, 1972.

outperform the usual one-point rule at the cost of only a slight increase in processing time. The applicability of such rules to satellite data processing remains to be investigated. We propose the following three-stage test of nine-point rules on LANDSAT data from the LACIE experiment.

The first stage will be an exploratory comparison of all the nine-point rules on two LACIE intensive study sites. Interior points of areas associated with ground truth will be identified. Half these fields will be designated as training sets and half as test sets. The measure of performance will be the percentage of pixels correctly classified computed separately for the training and test fields.

After considering the results from the first stage and from previous testing on aircraft data, no more than three nine-point rules will be chosen for extensive testing at the second stage. The procedure will be the same as for the first stage except that there will be fewer rules and more study sites tested.

The third stage will be to test the second-stage rules as acreage estimators in two ways. The first way will be to run the rules as classifiers and estimate the wheat acreage, say, by the number of pixels classified as wheat. (The number of pixels rather than the number of acres will measure area.) The second way will be to calculate the posterior probability of wheat for each pixel and sum the posterior probabilities.

The purpose of the three-stage design is to get the most information for the least computer running time. A detailed description of the test plan is given as Appendix D.

PART II
BOUNDARY DETECTION

James M. Gleason

6
INTRODUCTION

The ultimate success of operational remote sensing systems depends upon the speed and accuracy with which information can be extracted from the data generated by these systems. The farmer, the city planner, and the conservationist, among others, desire reliable, up-to-date information. They will utilize the system which provides this information in the most cost-effective manner. A number of factors contribute to the appeal of remote sensing systems as a feasible source of this information. Large ground areas can be surveyed rapidly and frequently from satellites or high altitude aircraft. Multispectral scanners can detect radiation from the ground in each of several wavelength bands and store this data in a form acceptable to a computer. High speed digital computers or specially designed hardware systems can rapidly process the tremendous bulk of data generated by these systems after some necessary, preliminary operations have been performed. Further advancements, however, must be achieved in reducing the amount of time required for these pre-processing operations. Also, more accurate, computerized methods for extracting the desired information from the recorded data must be developed. The cost-effectiveness of remote sensing systems cannot be firmly established until improvements are made in the accuracy and efficiency of these operations.

One application of remote sensing which has received considerable attention is the determination of major crop categories in an agricultural region. A multispectral scanner essentially decomposes the scene into a matrix of data points, or pixels (picture elements), each pixel corresponding to the ground resolution element size of the scanner. A multi-dimensional data vector is recorded for each pixel with the data value in each channel proportional to the

radiance reflected from the ground in a particular spectral band. A maximum likelihood recognition rule is often employed to classify each of these data points as one of the crop categories of interest[4]. This rule is quite easily implemented on the computer, with a quantity proportional to the likelihood function computed for each crop category. The data point is classified as the crop which maximizes the likelihood function. Before each pixel can be processed in this manner, however, a mean vector and covariance matrix must be estimated for each crop category. This preliminary operation is an error-prone and painstaking task which requires locating in the data a reasonable number of fields of each category and estimating the necessary parameters from these data points. Graymaps of several channels of data, ground truth maps and photographs of the area must be carefully examined to determine the location of these fields and the precise data points which correspond to each. The actual recognition processing is performed on each pixel individually, independent of all other points. Only the spectral characteristics of the crop categories are utilized in this decision process.

This investigation has focused upon methods for locating agricultural field boundaries as one likely way of improving the procedures just described. Field boundaries are an important feature contained in agricultural data sets. The development of efficient techniques for extracting this feature will permit its utilization as a means of reducing the time required for preliminary processing operations and also increasing recognition accuracy. A computerized boundary detection algorithm would alleviate much of the manual work required to locate fields so that the necessary parameters can be estimated. Classification accuracy could be increased by using spatial features as well as spectral features. The location of field boundaries would allow entire fields to be classified as a whole, reducing the effects of random variations within each field.

Field boundaries could be employed to increase accuracy and efficiency in other manners also. Time-consuming mixtures algorithms are used for estimating the proportion of each crop category which contributed to a pixel overlapping

[4] R.B. Crane, W. Richardson, and W.A. Malila, A Study of Techniques for Processing Multispectral Data, Technical Report 31650-155-T, ERIM, Ann Arbor, Michigan 1973.

more than one field[5]. If the location of field boundaries were known, these algorithms could be applied only in regions where mixtures will occur. Field boundaries could also be used for multi-temporal image registration. To make use of the time variations which occur among crop categories, data sets recorded at different times can be processed simultaneously. Field boundaries would provide an excellent feature to align these data sets because the boundaries are quite stationary over reasonable lengths of time.

The basic gradient and a number of modifications to it have been investigated as a means of boundary point detection. The detection of boundary points (not necessarily closed boundaries) is particularly useful for image registration and mixtures application. The basic thresholded gradient has been reported as an inherently "noisy" technique. However, the computational efficiency of the method is very appealing. Also, the possibility of improved performance from modifications to the basic technique, warranted a more thorough examination of the method.

The hypothesis testing technique for closed boundary formation, developed by LARS, Purdue [6,7] has been tested and evaluated. Closed boundaries are necessary for field identification and classification applications. Actually, two algorithms have been tested; one employing first-order statistics and the other also using second-order statistics. The evaluation of these algorithms was undertaken to gain a better insight into the closed boundary formation problem, and also to better determine the performance level of these methods.

Efforts have also been initiated to analyze the boundary detection problem in a detailed and sophisticated manner. The problem will be formulated in increasingly complex steps and each formulation will be analyzed using rigorous mathematical techniques. These formulations will allow the different atmospheric and system effects in the data to be isolated and thoroughly understood. Also, the possibility of employing different boundary features can be

[5] H.M. Horwitz, P.D. Hyde, W. Richardson, Improvements in Estimating Proportions of Objects from Multispectral Data, Technical Report 190100-25-T ERIM, Ann Arbor, Michigan 1974

[6] R. Kettig & D. Landgrebe, Automatic Boundary Finding and Sample Classification of Remotely Sensed Multispectral Data, LARS Information Note 041773, Purdue University.

[7] J.N. Gupta & P.A. Wintz, Closed Boundary Finding Feature Selection and Classification Approach to Multi-Image Modeling, LARS Information Note 062733, Purdue University.

investigated. Statistical signal detection and parameter estimation methods will be used to analyze each formulation. One basic formulation of the problem has been investigated in this manner.

7
GRADIENT TECHNIQUE AND MODIFICATIONS

The basic gradient technique and the modifications to it, which have been developed and tested, are boundary point detectors. These methods will result in certain pixels being designated as boundary points. They are not designed to guarantee that the boundaries which are detected will also be closed. As mentioned previously, some applications do not require closed boundaries. The removal of this closure constraint will, generally, simplify the detection problem. The gradient is basically a two-dimensional differentiation procedure. It can be used to indicate both the amount of contrast about a given point and the direction in which this contrast takes place.

The approximation to the spatial gradient used by Anuta[8] and also discussed by Rosenfeld[9] and Duda & Hart[10] is illustrated in Figure 8 for a 3 pixel by 3 pixel scene area.

A	B	C
D	E	F
G	H	I

$$DX = |(C+F+I)-(A+D+G)|$$
$$DY = |(A+B+C)-(G+H+I)|$$
$$MAG = DX + DY$$

FIGURE 8

APPROXIMATION TO SPATIAL GRADIENT

The DX and DY components indicate the magnitude of the contrast occurring over the 3 x 3 array in two orthogonal directions. They can be added as vectors to yield a geometric magnitude and a directional component for this contrast. A reasonable approximation to this geometric magnitude is formed by the sum of the DX and DY components (MAG).

The gradient operator is employed in a sliding-window mode, such that its magnitude (MAG) is computed at each point in the scene. Boundary points are defined as those points at which the magnitude exceeds a specified threshold value.

The following discussion refers to the 3 x 3 pixel array shown in Figure 8. The basic gradient and modifications which have been developed and tested are explained below.

1. BASIC GRADIENT

This is the gradient approximation mentioned previously with the addition of the geometric magnitude.

$$DX = |(C+F+I) - (A+D+G)|$$

$$DY = |(A+B+C) - (G+H+I)|$$

$$MAG = DX + DY$$

$$GMAG = \sqrt{DX^2 + DY^2}$$

2. MEDIAN GRADIENT

This is very similar to the basic gradient except that the median of three data values is used rather than their sum.

$$DX = |MEDIAN(C,F,I) - MEDIAN(A,D,G)|$$

$$DY = |MEDIAN(A,B,C) - MEDIAN(G,H,I)|$$

$$MAG = DX + DY$$

$$GMAG = \sqrt{DX^2 + DY^2}$$

3. ADJACENT GRADIENT

This technique uses adjacent rows and columns rather than every other row and column.

$$DX = |(C+F+I) - (B+E+H)|$$

$$DY = |(A+B+C) - (D+E+F)|$$

$$MAG = DX + DY$$

$$GMAG = \sqrt{DX^2 + DY^2}$$

4. SECOND GRADIENT

This processing method uses either the DX and DY components, the

MAG component, or the GMAG component and computes a second gradient exactly as in methods 1, 2, or 3 above,

Some additional techniques have also been employed which are not strictly gradient methods. These are described below.

1. AVERAGE

This algorithm substitutes the average of the data values over the 3 x 3 array of points as the value of the center point. The gradient functions can then operate on these averaged data values.

$$MAG = \frac{1}{9}(A+B+C+D+E+F+G+H+I)$$

2. WEIGHTED AVERAGE

This technique operates on the gradient values rather than the original data values. It emphasizes boundary points by weighting the middle row and middle column of the 3 x 3 gradient array. Essentially, it is a correlation between a function, which peaks along the middle of the 3 x 3 array and slopes down at the sides, and the output of one of the gradient operators. (Primes are used in the expressions below to indicate that these are gradient values and not data values.)

$$DX' = .5(C'+F'+I') + 2(B'+E'+H') + .5(A'+D'+G')$$
$$DY' = .5(A'+B'+C') + 2(D'+E'+F') + .5(G'+H'+I')$$
$$MAG' = DX' + DY'$$

3. LINEAR BOUNDARY

This algorithm is designed to detect boundary points which lie on a straight boundary line segment over a 3 x 3 pixel array. The boundary segments must lie in either a horizontal, vertical or diagonal orientation through the center of the array. A boundary in any one of these orientations will have higher gradient values at the points which lie directly on the boundary and smaller values at the points on both sides of the boundary. The algorithm is thus designed to detect points about which the highest gradient values lie along a straight line through the center of a 3 x 3 pixel array. This scheme does not require

that a threshold value be specified. It is based on the expected spatial characteristics of the boundaries and on the relative magnitudes of the gradients over a 3 x 3 region centered at each point in the scene.

Specifically, the algorithm indicates a boundary point wherever gradient values, along three pixels in a straight line through the center of the 3 x 3 array, are all greater than the average gradient value taken over all nine points of the array. For example, a vertical boundary point would be indicated if the gradients at pixels B, E and H were all greater than the average gradient value calculated from the points A through I.

The detection of agricultural field boundaries would be enhanced by the use of more than one channel of the multispectral data. It is entirely possible that boundaries which are clearly evident in one channel, may not be evident in another channel. Unfortunately, the gradient techniques are inherently one-channel operations. The basic gradient is an approximation to the two-dimensional derivative of a two-dimensional function, a definition which cannot be extended to more than one function.

There are, however, a number of ways in which the gradient techniques can be applied to multispectral data. With the exception of the LINEAR BOUNDARY method, all of the gradient techniques produce a resultant output magnitude value for one channel of data. To handle the multiple channel case, these values can simply be summed over all of the channels of interest. Alternatively, boundary points could be detected in each channel separately, with the final multiple channel output indicating a boundary point whenever one has been detected in at least one channel, or possibly two or more channels. Another possible method would be to take the largest gradient value over all of the channels, as the final magnitude value.

8
HYPOTHESIS TESTING TECHNIQUE

Two methods of closed boundary formation developed by LARS, Purdue have been programmed, tested and evaluated. The two methods are based on statistical hypothesis tests, one using first-order statistics only and the other also using second-order statistics. Small, square arrays of data points are combined into larger contiguous pixel groups by a field building algorithm, beginning at the top of the scene and proceeding downward through it. Ideally, the pixel groups will correspond to the agricultural fields present in the scene. The outer edge of each group forms a closed boundary. These groups are formed such that the data values of all of the pixels within each group are statistically similar in each channel. This statistical similarity is determined in one case by testing the null hypothesis that the means of two sets of samples are equal against the alternative hypothesis that they are not equal. This test employs only first-order statistics. An additional test on the magnitude of the variance of each sample is also included. In the other case, second-order statistics are used to also test the null hypothesis that the variance of each set of samples is equal, against the alternative hypothesis that they are not equal. Two samples are merged together into one larger group, when the null hypotheses described above are not rejected. The degree of similarity required not to reject these hypotheses is governed by a chosen confidence level for each test. A complete description of the field building algorithm and the two statistical similarity tests is given in Appendix E.

The hypothesis testing algorithms which have been programmed for use in this investigation were intended to be as nearly equivalent to the LARS algorithms as possible. The program now allows for one of five confidence levels (90%, 95%, 98%, 99%, 99.9%) to be chosen for the similarity test. Also, only one size is allowed for the sample array, 2 pixels by 2 pixels. One slight difference is known to exist in the manner in which the first row of samples is processed. However, this change only affects the first row and should have no significant affect on the over-all results of the method.

9

TEST AND EVALUATION OF GRADIENT AND HYPOTHESIS TESTING TECHNIQUES

The gradient and hypothesis testing techniques have been tested on both aircraft and satellite data sets. The primary concern of this investigation has been the actual detection of field boundaries and not their use in the applications mentioned previously. Unfortunately, this emphasis leaves no other method for evaluating the results of these tests than a purely subjective one. The ultimate evaluation of any boundary detection technique will depend on how accurately and efficiently it can be utilized in a practical application. However, it is believed that reliable conclusions can be reached by a thorough analysis of the methods and the results which have been obtained.

9.1 BASIC GRADIENT AND MODIFICATIONS

The gradient boundary detection techniques were tested using one channel of the Imperial Valley data set (3/12/69, Run 5, 5000 ft., .478-.508 μm). This data set was chosen because of its usefulness in previous experiments for separating good processing techniques from poor ones. For simplicity of analysis only the one channel case was considered. All of the gradient techniques, with the exception of the LINEAR BOUNDARY method, require that a threshold value be set which the gradient magnitude must exceed to be designated as a boundary point. However, the use of one threshold value is not sufficient to obtain meaningful comparisons between these different techniques. The range of output values from these methods will vary considerably because of the particular computations involved in each. A threshold value which produces reasonable results for one technique may produce very poor results for another. These poor results for the latter method are caused by an improper threshold setting and are not truly indicative of the capabilities of that technique. To avoid this difficulty, a different threshold setting was used for each method to be compared. The threshold value was chosen to produce a fixed percentage of the total number of data points as boundary points. The value for each particular technique was determined by generating a histogram of the output values and obtaining the proper threshold setting from that histogram.

The following qualitative results have been observed from the gradient techniques which have been investigated:

1. The gradient techniques are, indeed, "noisy" as indicated in the literature and as expected. As the threshold is lowered, more true boundary points are detected but more false boundary points are also detected. Many false detections are made throughout the scene when the threshold is lowered to a level at which the low contrast boundary points are correctly detected. At this same level, the high contrast boundaries are significantly widened. Many errors are also made near wide roads and other such inhomogeneous areas.

2. The MEDIAN GRADIENT is insensitive to noise spikes. The other techniques will most often incorrectly detect these spikes as boundary points.

3. The MEDIAN GRADIENT values have a smaller range than the BASIC and ADJACENT GRADIENT techniques because the DX and DY components are each differences between two data values. In the other methods the differences are taken between two sums of three data values.

The ADJACENT GRADIENT has a smaller range of values than the BASIC GRADIENT. This can be attributed to the gradual change in data values across the 3 x 3 array. The change across adjacent rows is not as great as between two rows separated by one row.

The increased range of the BASIC GRADIENT makes it slightly less sensitive to noise (excluding noise spikes) than the other two methods. In the MEDIAN and ADJACENT methods all of the data is more lumped together and it is more difficult to discriminate the boundary points.

4. There is no apparent increase in accuracy using the geometrical gradient magnitude (GMAG) as opposed to the algebraic magnitude (MAG). However, GMAG requires more computation time than MAG.

5. After averaging (AVERAGE) the entire data set, all of the gradient methods widen the true boundaries but do not detect as many false boundary points. The averaging effectively blurs the scene. Noise variations are not as great but variations due to actual boundaries are also less pronounced and more spread out.

6. The WEIGHTED AVERAGE tends by its formulation to emphasize the horizontal and vertical boundaries. It widens the actual boundaries and also forces them to line up in vertical and horizontal orientations. The averaging feature of this method also decreases the noise.

7. The second gradient methods do not yield a significant increase in accuracy over the first gradient methods. The SECOND GRADIENT does eliminate some of the interior points in regions which were entirely covered with first gradient boundary points. This is due to the second differential character of this method and its insensitivity to uniform changes. This method also widens the boundaries.

8. The advantage of the LINEAR BOUNDARY method is that an arbitrary setting of a threshold value is not necessary. Two problems arise, however. First, the algorithm tends to be sensitive to the row nature of the agricultural fields. Second, some boundary points are not detected because, in fact, the boundary line segments which they lie on are not straight over a 3 pixel by 3 pixel scene area.

9.2 HYPOTHESIS TESTING TECHNIQUES

The two hypothesis testing methods of closed boundary formation have been tested and evaluated using one aircraft data set and one satellite data set. The aircraft data set was taken from Mission 43M of the Corn Blight Watch Experiment, segment 204. Four channels of data were utilized (.46-.49 μm, .50-.54 μm, .54-.60 μm, .61-.70 μm). This data set was chosen because of the acceptable classification accuracy obtained from it during the Corn Blight experiment. The satellite data set was taken from the "San Francisco" ERTS frame (1003-18175). This data set was chosen because of the large agricultural fields present in the scene.

For the data sets which were tested, the method using only first-order statistics produced superior results to that which also used second-order statistics. Also, the results for the aircraft data set were significantly better than those for the satellite data set. The data was analyzed in groups of 2 x 2 arrays of pixels, and although the satellite data set was chosen because of its large fields, the number of pixels in each field was still quite small. Many of the 2 x 2 cells overlapped more than one field, resulting in a significant loss of detail.

An important parameter of these methods is the confidence level which must be chosen. As this level is lowered, the number of errors of commission (detecting boundaries which do not exist) should increase. At the same time, the number of errors of omission (not detecting boundaries which do exist) should decrease. For the data sets used in this investigation, decreasing from the highest confidence level to the lowest, resulted in a minimal decrease in the number of errors of omission, as compared to the increase in the number of errors of commission. The errors of omission were negligible even at the highest confidence level (99.9%). The errors of commission, however, were unacceptably high at the lower confidence levels (90%, 95%, 98%) and only became reasonable at the higher confidence levels (99%, 99.9%). The variations between the fields in these data sets were of sufficient magnitude that the true boundaries could be detected even at the highest confidence level. The variations within the fields, however, were also of sufficient magnitude that many false boundaries were detected at the lower confidence levels. Only at the higher confidence levels, could these within-field variations be rejected, with reasonable accuracy, as possible boundaries.

It appears that reasonably accurate results can be obtained with these methods if the fields in the scene contain a fairly large number of pixels and the within-field variations are significantly less than the between-field variations. If these conditions are satisfied, the confidence level can be adjusted to reject the within-field variations as possible boundaries, but still detect the between-field variations. The basic concept of the field building algorithm is a very good procedure for forming closed boundaries. The generally, small number of pixels in each field in satellite data sets and the unknown nature of the variations of the data limit the application of these methods.

10

DETAILED ANALYSIS OF BOUNDARY DETECTION PROBLEM

The boundary detection problem is currently being approached in a thorough and sophisticated manner. The investigations conducted into the use of the gradient and hypothesis testing techniques have indicated that such an approach is required to obtain an acceptable solution. This problem is made difficult by the complex factors which are involved in the acquisition of the data and also by the inherently variable nature of the boundaries. The most significant factors which affect the data are the atmosphere and the characteristics of the sensing system. The atmosphere absorbs radiation as it passes from the ground to the sensor and scatters radiation into the field of view of the sensor. The resolution of the scanning system and the sampling scheme which is employed are but two of the system factors which also affect the data. These atmospheric and system effects result in the boundaries being encoded in the data in a very complex manner. The variability of the boundaries is due to the many possible combinations of ground classes which they may separate, and the generally unknown spectral and spatial characteristics of these classes. This variability results in a lack of prominent features which can be used to discrimininate between the boundaries and the remainder of the scene.

The problem is now being approached by formulating it in increasingly complex steps and using rigorous mathematical methods for analyzing each formulation. The problem is first formulated in its most basic aspects. Additional factors which affect the solution will be considered in later formulations. This sequential approach will allow each factor to be isolated and clearly understood. Each factor can first be considered independently and later in combination with others. In this way, the complex atmospheric and sensor system effects which encode the boundaries in the data can be analyzed and methods of decoding this data developed.

The mathematical techniques which are being employed are those of statistical signal detection and parameter estimation. These techniques have proven

useful in the extraction of information from radar and communications signals. The variability of the boundaries will be taken into account by formulating the problem in terms of signals of known form and signals with unknown parameters. Formulations with small amounts of variability are considered first and more complex cases will be considered later. The solution of these formulations of the problem will require that these various signals be detected in the data and then unknown parameters estimated. By analyzing each formulation with these rigorous statistical tools, optimal or near optimal methods of performing these operations may be developed and their performance predicted.

Only one basic formulation of the problem has been investigated thus far. This formulation assumes that continuous data is obtained over a spatial interval in which a boundary is known to exist. The mean values of the data from the two ground class separated by the boundary are assumed to be known. The data is contaminated by zero-mean, additive, white Gaussian noise. This noise reflects the variations which occur in the data from random atmospheric and system effects and also from variability within the ground classes. The data must be processed to obtain an estimate of the unknown boundary location.

The data, or received signal $r(y)$, is recorded over the spatial interval $[-Y, Y]$. This received signal consists of a signal of known form $s(y)$, shifted by the boundary location parameter γ, plus the noise signal $n(y)$.

$$r(y) = s(y-\gamma) + n(y) \qquad \begin{matrix}(-Y \leq y \leq Y) \\ (-Y < \gamma < Y)\end{matrix}$$

The known signal $s(y)$ extends over the infinite interval and has a form similar to that shown in Figure 9. The mean values of the two ground classes contained within the interval are M_1 and M_2. The exact shape of the transition region between these two values is primarily determined by the aperture function of the sensing system. Specifying a value for the parameter γ shifts this known signal form to the right or left of the origin by an amount which depends on the magnitude and sign of the value. The true boundary location is γ_o. The estimate of the boundary location $\hat{\gamma}$, should be as close to this true value as possible.

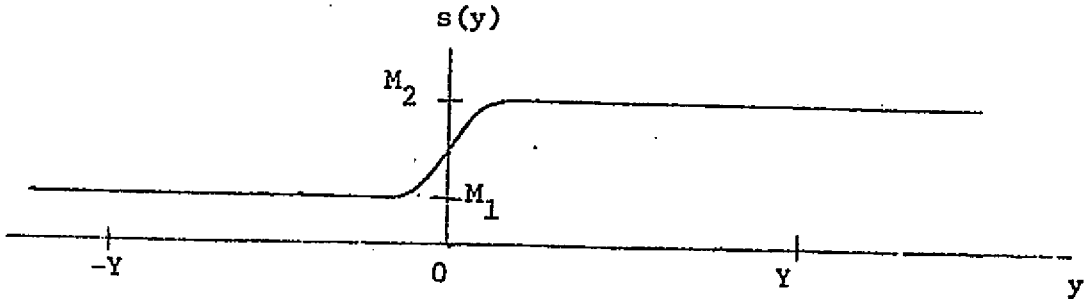


Figure 9
KNOWN SIGNAL FORM s(y)

Our estimation procedure has been to find the maximum likelihood estimate of the boundary location. Such estimates have been shown to be useful in many diverse applications and have desirable statistical properties. Estimation procedures which employ the prior probability of occurrence of the unknown parameter could also have been investigated. Frequently, however, the distribution of the parameter is rather flat, particularly in the region near the true value, and the resulting estimation procedure is equivalent to maximum likelihood. The maximum likelihood estimate of γ is that value $\hat{\gamma}$ which maximizes the likelihood function $p(r(y)|\gamma)$.

The log-likelihood function $\Lambda(\gamma)$, for continuous $r(y)$ on the interval $[-Y, Y]$, is given by

$$\Lambda(\gamma) = \ell n\ K - \frac{1}{N_o} \int_{-Y}^{Y} [r(y)-s(y-\gamma)]^2 dy$$

where K is a constant and N_o is the noise spectral density. The maximum likelihood estimate is that value $\hat{\gamma}$ which maximizes this expression when substituted for the parameter γ.

A necessary condition which the maximum likelihood estimate must satisfy, is that the partial derivative of the log-likelihood function must equal zero when evaluated at that value. Differentiating the expression for $\Lambda(\gamma)$ and evaluating it at $\gamma = \hat{\gamma}$, yields the result that

$$\int_{-Y}^{Y} [r(y)-s(y-\hat{\gamma})]\ s'(y-\hat{\gamma})dy = 0$$

must be satisfied for a maximum likelihood estimate.

The second part of this integral

$$\int_{-Y}^{Y} s(y-\hat{\gamma})\ s'(y-\hat{\gamma})dy$$

can be easily evaluated and is equal to

$$\frac{1}{2}[s^2(Y-\hat{\gamma})-s^2(-Y-\hat{\gamma})]$$

Thus, the estimate $\hat{\gamma}$ must satisfy the relationship

$$\int_{-Y}^{Y} r(y)\ s'(y-\hat{\gamma})dy = \frac{1}{2}\ [s^2(Y-\hat{\gamma})-s^2(-Y-\hat{\gamma})]$$

The right-hand side of this expression is essentially a bias term resulting from the non-zero values of the shifted signal at the endpoints of the observation interval. The left-hand side is obviously a correlation function, but the value of this expression can also be realized as a convolution operation. The output v(y) of a linear filter with impulse response h(y) and input r(y) defined on the interval [-Y, Y], is given by the convolution of h(y) and r(y).

$$v(y) = \int_{-Y}^{Y} r(u)\ h(y-u)\ du \qquad -Y \leq y \leq Y$$

Applying the received signal to a filter with impulse response

$$h(y) = s'(-y)$$

results in an output as a function of y which will be identical to the value of the left-hand side of the above expression as a function of the parameter $\hat{\gamma}$. The estimate $\hat{\gamma}$ is equal to the value of y at which the filter output is equal to $\frac{1}{2}\ [s^2(Y-\hat{\gamma})\ -\ s^2(-Y-\hat{\gamma})]$.

The distribution of the boundary location estimate obtained by this procedure can be derived in the high signal-to-noise ratio case. The derivation is shown in Appendix F. The maximum likelihood estimate $\hat{\gamma}$ has a Gaussian distribution with mean value γ_o and variance

$$V(\hat{\gamma}) = \frac{N_o}{2\int_{-Y}^{Y} [s'(y)]^2 dy}$$

The Cramer-Rao bound is also derived in Appendix F. This bound specifies the minimum variance which can be achieved by any estimate of the boundary location. In the high signal-to-noise ratio case, the variance of the maximum likelihood estimate is identical to the minimum variance specified by the Cramer-Rao bound. This estimation procedure yields not only an unbiased estimate but also a minimum variance estimate.

Consider, for example, a signal s(y) obtained as shown in Figure 10. The

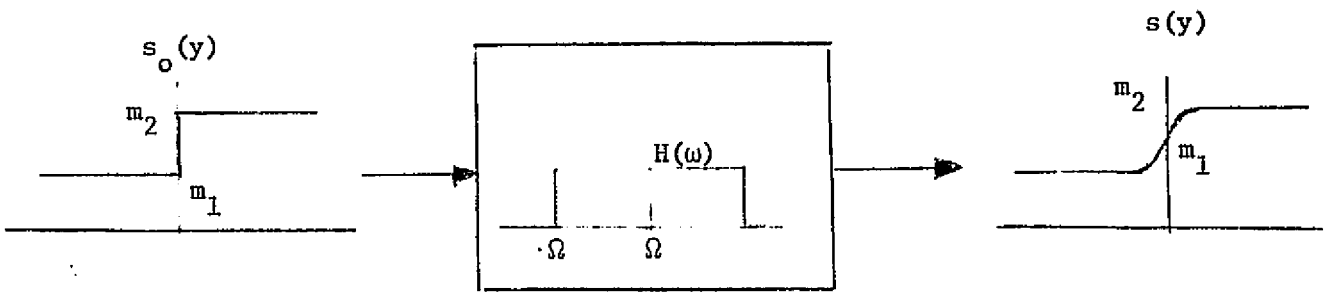


Figure 10
GENERATION OF SIGNAL s(y)

input signal $s_o(y)$ contains a perfect step function as its transition region. The signal is passed through the ideal low pass filter with transfer function $H(\omega)$. This filter bandlimits the output signal s(y) to frequencies less than Ω. The high frequency components resulting from the step function are removed by the filter, and the output signal s(y) is similar to that shown at the beginning of this section.

The variance of the maximum likelihood estimate of the boundary location for this signal s(y) will now be determined. The derivative of s(y) will have a value other than zero only in the immediate vicinity of y = 0. Letting

$$g(y) = s'(y)$$

and using Parseval's theorem,

$$\int_{-Y}^{Y} [s'(y)]^2 dy = \int_{-Y}^{Y} [g(y)]^2 dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega$$

where $G(\omega)$ is the Fourier transform of $g(y)$. The signal $s_o(y)$ has a derivative $s_o'(y)$ and Fourier transform $S(\omega)$ given by

$$s_o'(y) = (M_2 - M_1)\,\delta(y)$$

$$S(\omega) = M_2 - M_1 \qquad -\infty < \omega < \infty$$

The derivative of the output signal s(y) is equivalent to the output of the filter if the input were $s_o'(y)$. The Fourier transform of the derivative of the signal s(y) is, thus,

$$G(\omega) = M_2 - M_1 \qquad -\Omega < \omega < \Omega$$

The variance of the estimate $\hat{\gamma}$ is given by

$$V(\hat{\gamma}) = \frac{N_o}{2\int_{-Y}^{Y} [s'(y)]^2 dy} = \frac{N_o}{\frac{1}{\pi}\int_{-\infty}^{\infty} |G(\omega)|^2 d}$$

Substituting for the magnitude of $G(\omega)$ the variance of the maximum likelihood estimate is

$$V(\hat{\gamma}) = \frac{N_o \pi}{2\Omega(M_2 - M_1)^2}$$

The variance is proportional to the magnitude of the noise spectrum and inversely proportional to the bandwidth of the signal and the squared difference of the two mean values.

The analysis of this simplified formulation of the boundary detection problem is but a first step towards an acceptable solution. The results which have been presented, however, indicate the power of the method. The convolution procedure which has been developed for obtaining the maximum likelihood estimate is a useful operation in many applications. It can be implemented by approximating the required impulse response and utilizing either an electrical circuit or digital computer to perform the necessary filtering. The variance of the estimate obtained by this procedure is the minimum variance which can be obtained by any estimation procedure in the high signal-to-noise ratio case. In this sense, the estimate is optimal. The variance of the estimate for the bandlimited signal indicates the accuracy which can be expected in this case and the factors which affect this accuracy. If the values of these factors are known in advance, the performance of the technique can be predicted without having to rely on actual test results. Also, the factors which affect the estimate and their relative significance can be clearly determined and understood. The advantages of this approach will be more clearly evident as more complex formulations of the problem are considered.

11
CONCLUSIONS AND RECOMMENDATIONS

Investigations of the use of the gradient and hypothesis testing techniques have indicated the need for a thorough and sophisticated analysis of the boundary detection problem. The analysis is required because of the complex manner in which the boundaries are encoded in the data, and also because of the lack of prominent features by which the boundaries can be discriminated in the scene. The approach which has been adopted is to formulate the problem in increasingly complex steps and analyze each formulation using rigorous mathematical methods. In this manner, each factor which affects the encoding process can be systematically included in the formulation and clearly understood. Also, the use of various boundary features can be investigated by formulating the problem in terms of known signal forms to be detected and unknown parameter values to be estimated. Statistical signal detection and parameter estimation methods can then be employed to investigate the solution of each formulation. These methods will often result in the development of optimal or near-optimal solution techniques.

The basic formulation of the problem which has been analyzed indicates the power of the method. The solution technique which has been developed is easily implemented and yields a boundary location estimate which is unbiased and also optimal in the sense of minimum variance. The factors which affect the accuracy of this estimate for a certain case and their relative significance are also clearly indicated.

Additional formulations of the problem should be analyzed in a similar manner. These formulations should include the effects of the atmosphere, the resolution of the system and the sampling scheme. These formulations should also include assumptions which are less restrictive than those used in this simple formulation.

The results of the basic gradient technique and the modifications which were developed, were for the most part unsatisfactory. Test results demonstrated that, for a reasonable threshold setting, many true boundary

points were detected but too many false boundary points were also detected. In regions where large contrast boundaries existed, the gradient techniques significantly widened the boundaries. In regions of low contrast boundaries, many false boundary points were detected throughout the scene before the threshold was sufficiently lowered to the level that the low contrast boundary points were detected. The gradient methods are effective as an easily implemented means of boundary enhancement for visual examination. However, difficulties in choosing a proper threshold and a tendency to emphasize randomly occurring variations, must be overcome before these techniques can be effectively employed.

The results of the hypothesis testing methods are more difficult to categorize. The method using first-order statistics performed better than that which also used second-order statistics. The results achieved with the aircraft data set at the higher confidence levels were reasonable. The results for the satellite data set were less satisfactory.

The most significant aspect of these methods is the field building algorithm which guarantees closed boundaries. The results which can be achieved by these methods depend primarily on the within-field and between-field variations which occur in the data set. The confidence level determines the amount of within-field variation allowed. If the within-field variation is less than the between-field variation throughout the scene, good results can be achieved with a proper confidence level. As the within-field variations approach the between-field variations, however, the results will not be as satisfactory. Also, the small number of pixels in the fields contained within satellite data sets significantly limits the use of these techniques for satellite remote sensing operations.

APPENDIX A
DERIVATION OF BAYES9

The basic assumption of BAYES9 is that a pixel probably represents the same material as neighboring pixels. The following model expresses this assumption as a function of θ, a parameter between 0 and 1 measuring the degree of dependence between a pixel and its neighbor.*

Let I and J be neighboring pixels, let a and b be materials and let P stand for "probability that". Assume that there are prior probabilities P(I=a) for all materials a. Under conditions of complete independence

$$P(I=a|J=a) = P(I=a)$$
$$P(I=a|J=b) = P(I=a)$$

and under conditions of complete dependence

$$P(I=a|J=a) = 1$$
$$P(I=a|J=b) = 0$$

When the truth lies somewhere between dependence and independence, the model defines these probabilities to be a weighted average of the extreme values:

$$P(I=a|J=a) = (1-\theta)P(I=a) + \theta \cdot 1$$
$$P(I=a|J=b) = (1-\theta)P(I=a) + \theta \cdot 0$$

which, rewritten neatly, is

$$P(I=a|J=a) = (1-\theta)P(I=a) + \theta$$
$$P(I=a|J=b) = (1-\theta)P(I=a)$$

When $\theta = 0$, the neighboring pixels are completely independent. When $\theta = 1$, they are completely dependent. Thus θ can be thought of as a "coefficient of dependence" ranging between 0 and 1.

We will show that this definition satisfies the postulates of probability. Obviously P(I=a|J=a) and P(I=a|J=b) are between 0 and 1. It remains to show that

*The following ERIM personnel assisted in this derivation: J. Gleason originated the conditional probability definition of the model, R. Crane suggested an approach to the derivation, and R. Kauth corrected a significant error.

$$\sum_b P(J=b \text{ and } I=a) = P(I=a)$$

which is equivalent to showing that

$$\sum_b \frac{P(J=b \text{ and } I=a)}{P(I=a)} = 1$$

i.e., that

$$\sum_b P(J=b \mid I=a) = 1$$

by the definition of conditional probability. Now

$$\sum_b P(J=b \mid I=a) = P(J=a \mid I=a) + \sum_{b \neq a} P(J=b \mid I=a)$$

$$= (1-\theta)P(J=a) + \theta + (1-\theta) \sum_{b \neq a} P(J=b)$$

$$= (1-\theta) \sum_b P(J=b) + \theta$$

$$= 1 - \theta + \theta = 1$$

because $\sum_b P(J=b) = 1$. Q.E.D.

We will establish the relationship between θ, the number k of materials and the probability p that two neighboring pixels are the same material. This relationship holds when all the prior probabilities P(I=a) are equal, and since there are k of them,

$$P(I=a) = 1/k$$

for every material a. p, the probability that a pixel is the same material as its neighbor, is more precisely

$$P(I=a \mid J=a) = (1-\theta)P(I=a) + \theta$$

so

$$p = (1-\theta)\ 1/k + \theta$$

Alternatively,

$$\theta = \frac{kp-1}{k-1}$$

Let X and y be data values from I and J respectively. We will derive the probability that I=a given data X from I and data y from J:

$$P(I=a \mid X,y) = \frac{P(X \text{ and } y \text{ and } I=a)}{P(X \text{ and } y)}$$

by the definition of conditional probability. Let g be P(X and y). The probability we are seeking can be expanded as

$$\sum_b P(X \text{ and } y \text{ and } I=a \text{ and } J=b)/g$$

$$= \sum_b P(X \text{ and } y \mid I=a \text{ and } J=b)P(I=a \text{ and } J=b)/g$$

by the definition of conditional probability. Now we assume that once a and b are given, X and y are distributed independently, so the expression becomes

$$\sum_b P(X \mid I=a)P(y \mid J=b)P(I=a \text{ and } J=b)/g$$

where $P(X \mid I=a)$ is the density function of X given I=a.

The expression becomes

$$\frac{P(X \mid I=a)}{g}\left[P(y \mid J=a)P(I=a \text{ and } J=a) + \sum_{b \neq a} P(y \mid J=b)P(I=a \text{ and } J=b)\right]$$

By the θ model definitions

$$P(I=a \text{ and } J=b) = P(I=a \mid J=b)P(J=b)$$

$$= P(J=a)[(1-\theta)P(I=a)+\theta] \qquad \text{when } b = a$$

$$= P(J=b)[(1-\theta)P(I=a)] \qquad \text{when } b \neq a$$

We are assuming P(I=a)=P(J=a). So the expression is

$$\frac{P(X|I=a)}{g}\,[P(y|J=a)P(I=a)((1-\theta)P(J=a)+\theta)$$

$$+\sum_{b\neq a} P(J=b)(1-\theta)P(I=a)P(y|J=b)$$

$$=\frac{P(I=a)P(X|I=a)}{g}\,[\theta P(y|J=a)+(1-\theta)\sum_{b} P(J=b)P(y|J=b)].$$

This formula expresses the a posteriori probability that I=a given the data X and y from the neighboring pixels I and J. We will now extend this formula to a center pixel I_0 and eight neighboring pixels $I_1 \ldots I_8$ with data values X_0 and $X_1 \ldots X_8$, respectively. The derivation is analogous to the two-pixel case.

The decision rule will be to choose that material a for which the a posteriori probability of a, given the nine data values $X_0, X_1, \ldots, X_8$, is greatest. This probability is

$$P(I_0=a|X_0,\ldots,X_8) = P(X_0,\ldots,X_8 \text{ and } I_0=a)/g$$

where g stands for $P(X_0,\ldots,X_8)$, by the definition of conditional probability. This expression can be expanded to a sum of joint probabilities:

$$\sum_{b_1}\cdots\sum_{b_8} P(X_0,\ldots,X_8 \text{ and } I_0=a,\ I_1=b_1,\ldots,I_8=b_8)/g$$

$$=\sum_{b_1}\cdots\sum_{b_8} P(X_0,\ldots,X_8|I_0=a,\ldots,I_8=b_8)P(I_0=a,\ldots,I_8=b_8)/g \qquad (12)$$

We could, at this point, get bogged down with a complicated model for $P(I_0=a,\ldots, I_8=b_8)$, but a practical course to take is assume that the neighbors are independent of each other and then

$$P(I_1 = b_1,\ldots, I_8 = b_8 | I_0=a)=P(I_1=b_1|I_0=a)\ldots P(I_8=b_8|I_0=a) \qquad (13)$$

The rightmost factor of expression (12) can be put in this form by applying the definition of conditional probability:

$$P(I_0=a,\ldots,I_8=b_8) = P(I_1=b_1,\ldots,I_8=b_8|I_0=a)\ P(I_0=a)$$

$$=P(I_1=b_1|I_0=a)\ \ldots\ P(I_8=b_8|I_0=a)\ P(I_0=a)$$

by the assumption (13) just made. We continue to assume that the data $X_0,\ldots,X_8$ are distributed independently once the materials $b_0,\ldots,b_8$ are known, i.e., that

$$P(X_0,\ldots,X_8|I_0=a,\ldots,I_8=b_8) = P(X_0|I_0=a)\ldots P(X_8|I_8=b_8)$$

The desired a posteriori probability is

$$\sum_{b_1}\ldots\sum_{b_8} P(X_0|I_0=a)\ldots P(X_8|I_8=b_8)P(I_1=b_1|I_0=a)\ldots P(I_8=b_8|I_0=a)P(I_0=a)/g$$

$$= \frac{P(X_0|I_0=a)P(I_0=a)}{g} \left[\sum_{b_1} P(X_1|I_0=b_1)P(I_1=b_1|I_0=a)\right]\ldots\left[\sum_{b_8} P(X_8|I_8=b_8)P(I_8=b_8|I_0=a)\right]$$

which, written more compactly, becomes

$$= \frac{P(X_0|I_0=a)P(I_0=a)}{g} \prod_{i=1}^{8} \left[\sum_{b} P(X_i|I_i=b)P(I_i=b|I_0=a)\right]$$

Now

$$P(I_i=b|I_0=a)=(1-\theta)P(I_i=b) \text{ for } b \neq a$$

$$=(1-\theta)P(I_i=b)+ \theta \text{ for } b=a$$

We are assuming $P(I_i=b)$ is the same for all I_i and can be written $P(b)$. Thus the a posteriori probability is

$$\frac{P(X_0|I_0=a)P(a)}{g} \prod_{i=1}^{8} [\theta\, P(X_i|I_i=a)+(1-\theta) \sum_b P(b)\, P(X_i|I_i=b)]$$

The denominator g is the same for all materials. The decision is unaffected by its absence. When all the prior probabilities P(a) are equal to 1/k, then P(a) can be left out for the same reason. Also each square bracket term can be divided by θ. The resulting criterion is the one given in section 3.1.

$$P(X_0|I_0=a) \prod_{i=1}^{8} [P(X_i|I_i=a)+ \frac{1-\theta}{\theta} \frac{\sum_b P(X_i|I_i=b)}{k}] \qquad (1)$$

APPENDIX B
IMPLEMENTATION OF BAYES9

A multispectral software system at ERIM consists of subroutines PROCESS and POINT that mount, read and unpack data tapes, call modules to process the data, pack output data values, four to a word, and write an output tape. At the point-processing stage, each module accepts an input data point called DATUM consisting of NCHAN channel values, modifies the DATUM vector in some way, storing the output vector in DATUM. After all the prescribed modules have been called, POINT and PROCESS pack up DATUM, adding it on to the output line that will be written on tape. If several operations are to be performed, they can be done as separate jobs (with the intermediate tape for one job providing the input for the next) or they all can be run together (with each module picking up the output DATUM vector from the previous module). The modules are also called at an earlier stage when initial calculations are made, and at a later stage for final calculations and printing of results.

BAYES9 is a POINT module to carry out the nine-point Bayesian decision rule. Equal prior probabilities are assumed. This assumption is seldom departed from in multispectral processing, and the simpler formulation of BAYES9 that it permits runs faster than the general rule, uses less storage and avoids scaling problems. A later version of the module will allow the user to select either of the two rules.

The criterion used will be criterion (3) in section 3.1

$$P(X_0|a) \prod_{i=1}^{8} \left[\frac{P(X_i|a) + s + s \sum_b P(X_i|b)}{\varepsilon + s\varepsilon + s \sum_b P(X_i|b)} \right] \qquad (3)$$

where $s = (1-\theta)/\theta(k+1)$ and the summation is over the non-null densities. A preceding module, QRULE, calculates for each point X and each material a the exponent of the likelihood $P(X_o|a)$ and stores it as an integer C, with values

from 0 to 511, in DATUM. The density

$$P(X_o|a) = (2\pi)^{-\frac{k}{2}} e^{-\frac{1}{2}C}$$

is not computed by QRULE or used in the one-point decision rule because the exponent is sufficient for the decision. BAYES9 does need the true density, but the constant factor does not affect the decision so it is left out. $e^{-\frac{1}{2}C}$ is precalculated for all 512 possible values of C and stored in a floating point vector EXTAB. Thus the exponentiation, which would ordinarily be time-consuming, is carried out in a flash by referring to EXTAB(C).

In the nine-point processing modules, a subroutine SAVE9 saves the two most recent lines of unpacked data vectors. Before the saving takes place, the square bracket factors of expression (3) are calculated and stored in the saved line where DATUM would have ordinarily been stored. This allows each square bracket factor to be used eight times without being recomputed.

The module is most easily explained if I first describe the earlier, slower method of calculation and then show how it was modified. In this earlier method, the square bracket factors are multiplied together and then multiplied by $P(X_o|a)$ to get the decision criterion in floating point form. The material with the biggest criterion, corresponding to the largest a posteriori probability, is chosen. The number of the material chosen is put out as DATUM(1). The value of the criterion, appropriately scaled, is put out as DATUM(2).

This number is, for its crucial values, very close to zero and so is hard to scale. The transformation

$$-2 \log_e (\text{criterion})$$

makes it like a chi-square value, easily scaled. Let d stand for the criterion.

Because $\log_e(d)$ is a time-consuming library subroutine, the following speeded up procedure is implemented.

$$\log_e d = \log_2 d * \log_e 2$$

$\text{Log}_e 2 = .69$. Hence

$$-2 \log_e d = - \log_2 d * 1.3862944$$

The multiplication by 1.39 is not essential. The purpose of the transformation was to scale the criterion and this purpose is as well accomplished if the 1.39 is left out. Therefore $- \log_2 d$ is obtained from the floating point representation of d in the computer by the following rapid procedure.

The floating point representation of d in the IBM 7094 consists of a sign bit, an 8-bit exponent E and a 27-bit fraction M. $E-200_8$ is the power of 2 which multiplies the fraction to get d:

$$d = M * 2^{(E-200_8)}$$

$$-\log_2 d = -\log_2 M - (E-200_8)$$

$$= -\log_2 M + (200_8 - E)$$

We want $-\log_2 d$ to be an integer, but $-\log_2 M$ is a fraction filling in the gaps between the integer values of (200_8-E). Therefore we shift each term of $-\log_2 d$ left nine places (i.e., multiply by 512). The second term could be calculated by subtracting E from 200_8 and shifting left 9 bits. But it is even faster to precompute this result as a 512-valued table LOG2E and obtain it by a reference to the table. The first term, shifted, is also obtained from a 512-valued precomputed table LOG2M whose domain is the first nine bits of M. So to get $-\log_2 d$ we just compute

$$\text{LOG2M(M)} + \text{LOG2E(E)}$$

and shift right nine bits. Then 200 is added to make sure the criterion falls in the range 0...511.

This algorithm for computing the criterion requires eight floating point multiplies in the inner loop, which is repeated for each material at each pixel. A modified algorithm comes out with eight integer adds in the inner loop, a faster operation, and is therefore used in the current version of BAYES9.

For this algorithm, each square bracket quantity is subjected to the $-\log_2$ transformation before storing the values in the SAVE9 array. The final right shift is not performed at this time. When the product in criterion (3) is actually calculated, it is the sum of eight integers. The leftmost factor $P(X_o|a)$ is included in the calculation by one more integer add. The original form of $P(X_o|a)$ is $-2\log_e P(X_o|a)$ expressed as an integer between 0 and 511. A table LOGTAB(I) has precomputed I shifted left nine places and divided by 1.3862944 to get it in the form $-\log_2 P(X_o|a) * 512$ like the square bracket factors.

The number of $\log_2$ operations performed is one per pixel, which is the same as before. The line of center values is saved in a special vector because only one line needs to be saved and it would be a waste of space to save two lines of this data in SAVE9. The criterion is shifted right nine places before being put out as DATUM(2), so that it will fall in the range 0...511.

The computation of the criterion requires many operations, none of them lengthy by themselves. In this situation, much time is saved by programming the point-processing routine in assembly language. Such a version was written and found to give results identical to those of the MAD* version. A timing run was made using six out of ten channels, 200 points per line and 60 lines with the following results:

QRULE alone	203.9"
QRULE + BAYES9 (MAD)	265.4"
QRULE + BAYES9 (assembly)	225.4"

Subtracting the QRULE alone time from the other two, we find that the additional time taken by the BAYES9 module was, in each case

BAYES9 (MAD)	61.5"
BAYES9 (assembler)	21.5"

*a source language similar to ALGOL

so the writing of the module in assembly language cut the running time for the BAYES9 module by a factor of 3. The assembly-language version of QRULE + BAYES9 took only 11% longer than the one-point rule QRULE alone.

The left side of the null test has been scaled by the $-\log_2$ transformation and so the right side becomes $-\log_2\varepsilon_2 + 200$. To estimate $-\log_2\varepsilon_2$, we choose a value EXPLIM of chi-square representing a reasonable dividing line between points inside and outside the distribution, estimate a typical value $|R|$ of the determinant of a covariance matrix and compute

$$-\log_2 \varepsilon_2 = \frac{\text{EXPLIM} + \log_e |R|}{1.38}$$

A first guess for EXPLIM is the value in the table of the chi-square distribution with the row corresponding to the number of original data channels and the column corresponding to the .001 significance level.

We will now derive bounds for criterion(3). Each square bracket term can be written

$$\frac{P(X_i|a) + S\varepsilon + S \sum_b P(X_i|b)}{\varepsilon + S\varepsilon + S \sum_b P(X_i|b)} \tag{14}$$

where the sum is over the non-null materials. We abbreviate expression (14) as

$$\frac{P + S\varepsilon + S\Sigma}{\varepsilon + S\varepsilon + S\Sigma} \tag{15}$$

for ease of discussion. All the terms are > 0 and $P < \Sigma$. (15) can get small only when $P < \varepsilon$. The lower bound occurs when $P \to 0$. Then (15) is

$$\frac{S\varepsilon + S\Sigma}{\varepsilon + S\varepsilon + S\Sigma} \tag{16}$$

The bigger Σ is, the closer (16) is to 1. Hence the lower bound occurs when $\Sigma \to 0$ and is therefore

$$\frac{S\varepsilon}{\varepsilon + S\varepsilon} = \frac{S}{1+S}$$

Expression (15) increases from 1 as P increases from ε. Hence

$$1 < \frac{P + S\varepsilon + S\Sigma}{\varepsilon + S\varepsilon + S\Sigma} < \frac{P + S\varepsilon + S\Sigma}{S\varepsilon + S\Sigma} < \frac{P + S\Sigma}{S\Sigma}$$

because the effect of adding $S\varepsilon$, top and bottom, is to bring the fraction closer to 1.

$$\frac{P + S\Sigma}{S\Sigma} < \frac{\Sigma + S\Sigma}{S\Sigma} = \frac{1 + S}{S}$$

because $P < \Sigma$. Thus we have shown

$$\frac{S}{1 + S} < \text{square bracket factor of (3)} < \frac{1 + S}{S} .$$

Now $S = \frac{1-\theta}{\theta(k+1)}$ $\quad 1 + S = \frac{\theta k + \theta + 1 - \theta}{\theta(k+1)} = \frac{\theta k + 1}{\theta(k+1)}$

$$\frac{S}{1 + S} = \frac{1 - \theta}{\theta k + 1}$$

So

$$\frac{1-\theta}{\theta k+1} < \text{square bracket factor of (3)} < \frac{\theta k + 1}{1-\theta}$$

Hence

$$P(X_o|a) \left(\frac{1-\theta}{\theta k+1}\right)^8 < \text{expression (3)} < P(X_o|a) \left(\frac{\theta k+1}{1-\theta}\right)^8$$

The bounds show that no square bracket factor in criterion(3) is an unreasonably large or small number. LOG2E therefore does not have to be dimensioned higher than $300_8 = 192_{10}$.

The development of a criterion analogous to (3) when the prior probabilities are unequal is as follows. We suppose that each defined material b has a positive prior probability P(b), that the null class N has a positive prior probability P(N) and that

$$\sum_{\substack{\text{all } b \\ \text{including } N}} P(b) = 1.$$

The BAYES9 criterion in this case (given in Appendix A) is

$$\frac{P(X_o|a)\ P(a)}{g} \prod_{i=1}^{8} [\theta P(X_i|a) + (1-\theta) \sum_b P(b)\ P(X_i|b)] \tag{17}$$

where the summation is over all materials and null. Criterion(17) for the null class is

$$\frac{\varepsilon\ P(N)}{g} \prod_{i=1}^{8} [\theta\ \varepsilon + (1-\theta) \sum_b P(b)\ P(X_i|b)] \tag{18}$$

We choose null when

$$\frac{\text{criterion(17)}}{\text{criterion(18)}} < 1$$

i.e., when

$$\frac{P(X_o|a)\ P(a)}{P(N)} \prod_{i=1}^{8} \left[\frac{\theta P(X_i|a) + (1-\theta) \sum_b P(b)P(X_i|b)}{\theta\varepsilon + (1-\theta) \sum_b P(b)P(X_i|b)} \right] < \varepsilon_2$$

as in the derivation of criterion(3). Simplifying a little, we arrive at the criterion

$$P(X_o|a)\ Q(a) \prod_{i=1}^{8} \left[\frac{P(X_i|a) + t \sum_b P(b)P(X_i|b)}{\varepsilon + t \sum_b P(b)P(X_i|b)} \right]. \tag{19}$$

where $Q(a) = P(a)/P(N)$ and $t = (1-\theta)/\theta$. $-\log_2$ criterion(19) is tested against $-\log_2 \varepsilon_2$ when a map is made or when the recognitions of each material are counted up.

APPENDIX C

IMPLEMENTATION OF PRIOR9 AND PREF9

In Sections 3.2 and 3.3, we derived two new nine point rules, PREF9 and PRIOR9. The classification criterion for PREF9 with respect to each material a is the answer to the question "Given the data from the nine pixels, what is the probability that if we choose a pixel at random from the neighborhood it is material a?" This criterion written in mathematical symbols is

$$\frac{1}{9} P(a|X_o) + \ldots + \frac{1}{9} P(a|X_8)$$

where $X_o, \ldots, X_8$ are the data vectors from the center pixel and the eight neighboring pixels, respectively. The factor 1/9 is the same for all materials and can bo omitted. By Bayes formula, the PREF9 criterion is

$$\frac{P(a)P(X_o|a)}{\sum_b P(b)P(X_o|b)} + \ldots + \frac{P(a)P(X_8|a)}{\sum_b P(b)P(X_8|b)} \qquad (10)$$

In summary, the PREF9 criterion for material a is the posterior probability of material a averaged over the nine data values.

PRIOR9 is a rule that uses these estimates of the probabilities of the different material as the prior probabilities for a usual Bayesian decision on the center pixel. This rule carries out the principle that the prior probability that a pixel represents a given material is dependent on the neighborhood in which the pixel is located. The criterion for PRIOR9 is criterion (1) multiplied by the center likelihood $P(X_o|a)$:

$$P(X_o\ a)\ \left[\frac{P(a)P(X_o|a)}{\sum_b P(b)P(X_o|b)} + \ldots + \frac{P(a)P(X_8|a)}{\sum_b P(b)P(X_8|b)}\right. \qquad (5)$$

It was shown in section 3.2 that a reasonable null test can be derived by assuming that a null catagory N has a flat distribution of height ε. N is chosen when its criterion (5) is larger than that of the winning material The test turns out to be equivalent to choosing N if

$$\frac{P(X_o|a)\ \dfrac{P(X_o|a)}{\varepsilon+\sum_b P(X_o|b)} + \ldots + \dfrac{P(X_8|a)}{\varepsilon+\sum_b P(X_8|b)}}{\dfrac{1}{\varepsilon+\sum_b P(X_o|b)} \qquad \dfrac{1}{\varepsilon+\sum_b P(X_8|b)}} < \varepsilon_2^2 \qquad (9)$$

The ε on the right side of (9) is called ε_2 to remind us that it can be changed after PRIOR9 is run whereas the ε on the left side must be specified before that run.

Two strategies for calculating criterion (9) present themselves, The first is to store

$$p_{ia} = \frac{P(X_i|a)}{\varepsilon+\sum_b P(X_i|b)}$$

as an integer for each pixel and each material a and

$$p_{i\varepsilon} = \frac{\varepsilon}{\varepsilon+\sum_b P(X_i|b)}$$

as in integer in channel k+1, Criterion (9) is

$$\frac{p_{oa}\ \sum_{i=0}^{8} p_{ia}}{\dfrac{p_{o\varepsilon}}{\varepsilon}\ \dfrac{\sum_{i=0}^{8} p_{i\varepsilon}}{\varepsilon}} < \varepsilon_2^2$$

i.e.

$$\frac{p_{oa} \sum_{i=0}^{8} p_{ia}}{p_{o\varepsilon} \sum_{i=0}^{8} p_{i\varepsilon}} < \frac{\varepsilon_2^2}{\varepsilon^2}$$

Only the numerator would be calculated to determine the winning material (because the denominator is the same for all materials and therefore doesn't affect that decision) and then the denominator divided to perform the null test.

The second strategy uses the relationship

$$\sum_b p_{ib} + p_{i\varepsilon} = 1$$

(i.e., the posterior probabilities sum to 1.) Hence

$$\sum_{i=0}^{8} p_{i\varepsilon} = 9 - \sum_b \sum_{i=0}^{8} p_{ib}$$

Hence

$$\frac{1}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{1}{\varepsilon + \sum_b P(X_8|b)} = \frac{9 - \sum_b \sum_{i=0}^{8} p_{ib}}{\varepsilon}$$

Hence criterion (9) is

$$\frac{P(X_o|a) \sum_{i=0}^{8} p_{ia}}{9 - \sum_b \sum_{i=0}^{8} p_{ib}} > \frac{\varepsilon_2^2}{\varepsilon}$$

Saving channel k+1 for $p_{i\varepsilon}$ is no longer necessary, but now a line of center pixel vectors $P(X_o|a)$ must also be saved. The tradeoff in storage is between saving two lines with one extra channel vs saving one extra line of k channels. Timewise, the balance is between summing 9 integers $p_{i\varepsilon}$ and summing k integers $\sum_{i=0}^{8} p_{ib}$ and subtracting from 9. There is also retrieval bookkeeping in getting at the values $p_{i\varepsilon}$ in the first strategy.

A modification of the second strategy is most economical of storage and not costly of time.

$$p_{ia} = \frac{P(X_i|a)}{\varepsilon + \sum_{b} P(X_i|b)}$$

is saved as an integer in k channels of every pixel in two lines.

$$f_{o\varepsilon} = \varepsilon + \sum_{b} P(X_i|b)$$

is saved in floating point for each pixel of a single line. The chosen material is the one minimizing

$$p_{oa} \sum_{i=0}^{8} p_{ia} \tag{20}$$

as before. Because this criterion is mostly a sum of integers (a very rapid operation on the computer)and because any decision criterion is calculated for every material a at every pixel, this criterion takes relatively little time.

The null test criterion (9) is calculated in the form

$$\frac{f_{o\varepsilon}\left[p_{oa} \sum_{i=0}^{8} p_{ia}\right]}{9 - \sum_{b} \sum_{i=0}^{8} p_{ib}} \tag{21}$$

for the winning material only. The square bracket factor in the numerator is the winning criterion (20). The sum $\sum_{i=0}^{8} p_{ib}$ is calculated for the criterion (20) of every material, so it is simple to keep the running total

$$\sum_{b} \sum_{i=0}^{8} p_{ib}$$

The square bracket term and the denominator are each converted to floating point and then the floating point expression (21) is calculated.

In order to express p_{ib} as an integer, the denominator of p_{ib}

$$\varepsilon + \sum_{b} P(X_i|b)$$

is divided by 100000 and then stored as $f_{o\varepsilon}$. This has the effect of multiplying p_{ib} by 100000. The denominator of (21) becomes

$$900000 - \sum_{b} \sum_{i=0}^{8} p_{ib}$$

The scaling cancels out for every factor in (21) but p_{oa}. When p_{oa} is multiplied by the reverse-scaled $f_{o\varepsilon}$, the original scale is restored.

In order that the null test can be performed at map-making time, the null test criterion is scaled by the transformation

$$-4 \log_2 [\text{criterion (21)}] + 100$$

and stored as an integer in an output channel. The transformation is quickly carried out by exactly the same method as was described in Appendix B.

For PREF9, the criterion for deciding among materials is the averaged posterior probability of material a

$$\sum_{i=0}^{8} p_{ia}$$

(the factor 1/9 has been omitted because it is the same for all materials.) The null test is

$$\frac{\sum_{i=0}^{8} P_{ia}}{9 - \sum_{b} \sum_{i} P_{ib}} < 1 \qquad (22)$$

and is scaled by the transformation

$$-8 \log_2 [\text{criterion (22)}] + 200$$

Because PRIOR9 requires the calculation of expressions used in PREF9, both rules are carried out in the same module containing a control variable that can be set to activate either or both of the two rules. If either rule is used alone, the number of the winning material is put out as channel 1 and the transformed null test criterion as channel 2. If both rules are requested, PRIOR9 puts out channels 1 and 2, PREF9, channels 3 and 4 and QRULE, the usual one-point decision rule, channels 5 and 6.

The input to the processing module PRIOR9 is the output of the one-point decision rule module QRULE, which can be run to put out $-2 \log_e$ (density of material b) for each material b in channels 3 through k+2, the smallest of these numbers (corresponding to the largest density) in channel 2 and the number of the material with the largest density (i.e., the one-point rule choice) in channel 1. For each input number y_1 corresponding to a material b, it is necessary to calculate $P(X|b) = e^{-\frac{1}{2}y}$. Because y is an integer between 0 and 511, this calculation is very rapidly accomplished by referring to the y^{th} element of a precalculated vector of length 512. These remarks about input also apply to the module BAYES9 (see Appendix B.)

We thought that the user would select both PREF9 and PRIOR9 only when the principal goal was experimental comparison of the two rules, and in that case, it would be convenient to have the QRULE results also on the same tape to minimize the number of tape mounts in map-making and to allow for the computation of agreements and disagreements among the three rules; hence the passing along of the QRULE results into channels 5 and 6.

The division of roles between ε and ε_2 in criterion (9) could have been made another way, namely

$$\frac{P(X_o|a)\left[\frac{P(X_o|a)}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{P(X_8|a)}{\varepsilon + \sum_b P(X_8|b)}\right]}{\frac{\varepsilon}{\varepsilon + \sum_b P(X_o|b)} + \ldots + \frac{\varepsilon}{\varepsilon + \sum_b P(X_8|b)}} < \varepsilon_2 \tag{23}$$

We will now show that this different division of roles results in the same numerical procedures and hence effectively the same null test as previously derived.

Following the first implementation strategy, test (23) can be written

$$\frac{P_{oa} \sum_{i=0}^{8} P_{ia}}{\frac{P_{o\varepsilon}}{\varepsilon} \sum_{i=0}^{8} P_{i\varepsilon}} < \varepsilon_2$$

i.e.,

$$\frac{P_{oa} \sum_{i=0}^{8} P_{ia}}{P_{o\varepsilon} \sum_{i=0}^{8} P_{i\varepsilon}} < \frac{\varepsilon_2}{\varepsilon}$$

which is the same as before except that the right side was $\varepsilon_2^2/\varepsilon^2$. If ε is set theoretically and not empirically, then $\varepsilon_2 = \varepsilon$ and both tests are identical. If ε_2 is set empirically then the right side is set empirically and it makes no difference what it is. In other words, the right side is set to make a good looking map or to leave unclassified a certain respectable percentage of points, so all that matters is what null test criterion is stored in output channel 2, and that is the same for both formulations of the null test.

Following the second implementation stragety, test (23) is

$$\frac{P(X_o|a) \sum_{i=0}^{8} p_{ia}}{9 - \sum_{b} \sum_{i=0}^{8} p_{ib}} < \varepsilon_2$$

which is the same as before except that the right side was $\varepsilon_2^2/\varepsilon$ and the remarks made about the comparison of first strategy tests apply.

We tried PRIOR9 and PREF9 run jointly with QRULE on 6-channel data (as a subset of 10 original channels) from 60 lines with 200 points per line. The run took only 12% longer than QRULE run alone. (This compares with 11% longer for BAYES9 and 16% longer for the outmoded BAYES8). No significant saving of time would result from running PRIOR9 or PREF9 alone because all the most time-consuming calculations would still have to be made.

APPENDIX D

A PLAN TO TEST AND EVALUATE NINE-POINT RULES ON SATELLITE DATA

The purpose of the test is to compare the performance of nine-point classification rules on satellite data. The test will be carried out on data from the LACIE intensive study sites. It will determine which rules best recognize wheat, whether their recognition is better than that of the usual one-point rule QRULE, and if so, how much better.

To economize on computer running time, the test will be carried out in three stages. The first stage will be an exploratory comparison of all the rules on two LACIE intensive study sites. This exploratory stage is necessary because the performance of nine-point rules on LANDSAT data has not previously been tested. After considering the results from the first stage and from previous testing on aircraft data, we will choose no more than three nine-point rules for extensive testing in the second stage. The third stage will be to test the second-stage rules as acreage estimators.

In the first stage, an attempt will be made to find two LACIE 3 x 3 mile intensive study sites on which the usual classification rule QRULE does not give perfect results, so as to allow room for improvement. Pixels that are clearly interior points of fields will be identified; these are the pixels for which we are confident of the ground truth. Half of these field interiors will be designated as training sets and the rest as test sets. The training sets will be clustered and combined into a small number of signatures for each material. This set of signatures will be called SIGS 1. The same data set will be used twice by designating the former test sets as training sets and the training sets as test sets, clustering the new training sets and combining them into a set of signatures SIGS 2. To preserve resolution, the data will not be rotated.

The data from the 73 x 105 pixel rectangle enclosing the intensive study site will be processed by QRULE, using first SIGS 1 then SIGS 2. The result is an output tape with two files, corresponding to SIGS 1 and SIGS 2, respectively, with the QRULE choice in channel 1, the QRULE null test

criterion in channel 2, and in the succeeding channels, a normal density for each material. These densities are input for all the nine-point rules. The processing time for such a small rectangle is negligible compared to the time for identifying the pixels interior to the fields.

The QRULE results will be tallied by running the processing module TALLY on the QRULE output tape. if the results are so good that they allow little room for improvement, then it would be a waste of time to test whether other decision rules improve performance. Therefore either the signatures would be degraded by combining them into one signature for each material or another intensive study site or the same one at a different time of year would be chosen.

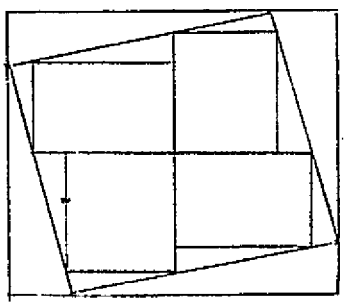
If the QRULE results allow room for improvement, then the enclosing rectangle will be processed by the modules LIKE9 with m=9, AVE9 with t=1, VOTE9, PREF9, PRIOR9, and BAYES9 with θ = .1, .3, .5, and .7. By storing the programs in separate core loads, this processing can all be done in one job with the QRULE output tape as input and an output tape with 18 files of pixel identifications, two files (for SIGS 1 and SIGS 2) per rule. The module TALLY will be run on this processing output tape to count the number of identifications of each material in each field interior and punch these numbers on cards. The cards will be input to a program DISPLAY to average the results and display them in tabular form. The results will be given separately for training and test fields, and in another breakdown, for large fields and small. The measure of performance displayed will be the percent misclassified. The total processing time will not be lengthy because the enclosing rectangle (73 x 105 pixels) is so small.

The procedure just described is applied to two intensive study sites Then we will look at the results from the two sites and, with less attention, to previous aircraft results to decide which rules are promising enough to merit more lengthy testing. Not more than three will be chosen.

The procedure for the second stage will be the same as for the first except that there will be more sites tested and fewer rules. We would attempt

to run tests on enough sites to determine whether the best nine-point rules significantly outperform QRULE in the wheat study and estimate how much improvement is to be expected. If, as seems likely, BAYES9 is one of the finalists, we will try to estimate an optimal value or range of values for the parameter θ.

In the third stage we compare the performance of the nine-point rule finalists and QRULE as acreage estimators. The pixels of the site will be assigned to four rectangles as shown:



The rectangles will be located on photographic enlargements and the wheat area measured by planimeter for each rectangle. This area will be divided by the area of the rectangle and then multiplied by the number of pixels in the rectangle to obtain the true wheat acreage (except that number of pixels is used as a measure of area rather than acres). QRULE and the nine-point rule finalists will be compared as acreage estimators in two ways. The first way will be for the rules to classify each pixel and count the wheat pixels in the rectangle. The second way will be to estimate the posterior probability of wheat for each pixel and sum these probabilities over the rectangle.

The results will be compared on the same site data used in the first two stages so that differing relative performances of the rules as classifiers and acreage estimators can be noted. Some reprogramming of the decision rule modules to function as acreage estimators will be needed.

APPENDIX E
DESCRIPTION OF HYPOTHESIS TESTING TECHNIQUES

A. Field Building Algorithm

The scene to be processed is first decomposed into a set of samples, each of which is a square array of points. Beginning with the first row of samples, the statistical similarity of adjacent samples is tested. Two samples are combined to form a pixel group when they are determined to be statistically similar. This group is then treated as a new, larger sample and is tested against the next sequentially occurring sample. A sample which is not similar to the pixel group to which the previous sample belongs is assigned to a new pixel group. The entire first row of samples is thus combined into a series of pixel groups, the samples in each group having been determined to be statistically similar. A group may contain only one sample or it may contain many samples.

Succeeding rows of samples, beginning with the second row, are processed by testing each sample against the pixel group which has, as one of its members, the sample in the preceding row directly corresponding to the test sample. Samples in two consecutive rows are directly corresponding when they are both displaced by the same number of samples from the beginning of each row. If the sample is not similar to this group and is not the first sample in the row, it is tested against the pixel group to which the preceding sample in the row belongs, if that sample has been assigned to a pixel group. If the sample passes the statistical similarity test, it is assigned to that group. If the test is not satisfied or if the preceding sample has not been assigned to a pixel group, the current sample is not assigned to any group and processing continues to the next sample. If the sample is similar to the pixel group to which the corresponding sample in the previous row belongs, it is assigned to that pixel group. Also, if the sample preceding the current sample has not been assigned to a pixel group, this preceding sample is tested against the group to which the current sample was just assigned. If they are not similar, the preceding sample remains unassigned, and processing continues to the next sample

following the current sample. If the preceding sample is similar, however, it is added to the pixel group to which the current sample was just assigned. This testing of preceding samples continues back along the row of samples until either a sample is encountered which is not similar to the group to which the current sample was assigned, or a sample is encountered which has already been assigned to a pixel group, or there are no more remaining samples to be tested.

Each row of samples is processed in this manner until the last sample in the row has been processed. At this point, the current row of samples is scanned back along the row until a sample is encountered which has not been assigned to a pixel group. The conditions under which unassigned samples may exist were explained previously. When an unassigned sample is encountered, it is assigned as the first element of a new pixel group. If the next sample moving back along the row is also unassigned, it is tested against the just formed pixel group. If it is similar to this group it is assigned to it. If it is not similar, it is used to initiate another new pixel group. Thus, whenever a consecutive series of unassigned samples is encountered, the first sample of the series is either assigned to the pixel group to which the preceding sample moving back along the row was assigned or it is used to initiate a new pixel group. After processing a row of samples in both the forward and reverse directions, all of the samples in the row have been assigned to a pixel group.

Once the entire scene has been processed in this manner all of the pixels have been assigned to contiguous sets of pixel groups. Ideally, each such set will correspond to one agricultural field. A closed field boundary is formed around each of these groups by those pixels which are part of the group but of which at least one of their eight immediate neighbors belongs to another group. Essentially, the closed boundaries form the outline in the scene of each of the pixel groups.

B. First-Order Similarity Test

Two statistical tests have been used by LARS to test the similarity between the multispectral data values of a sample and a pixel group. One

C-2

test uses first order statistics only, to test the hypothesis that the means of both are equal. Also included in this test in an empirical check on the variances of both. For one channel of data, the test proceeds as follows.

Let $(X_1,\ldots,X_N)$ be the one-channel data values of the points in the sample. Also, let $(Y_1,\ldots Y_P)$ be the data values for the points in the pixel group. Here, N is the number of pixels in the sample and P is the number of pixels in the group. The mean of the sample, M_X, and the mean of the group, M_Y, are given by:

$$M_X = \frac{1}{N} \sum_{i-1}^{N} X_i$$

$$M_Y = \frac{1}{P} \sum_{i=1}^{P} X_i$$

The sum of the sample values squared, Q_X, and the sum of the group values squared, Q_Y, are given by:

$$Q_X = \sum_{i=1}^{N} (X_i)^2$$

$$Q_Y = \sum_{i=1}^{P} (Y_i)^2$$

The normalized sum of squares of the sample values, NSS_X, and the normalized sum of squares of the group values, NSS_Y, are given by:

$$NSS_X = \sum_{i=1}^{N} (X_i - M_X)^2 = Q_X - N(M_X)^2$$

$$NSS_Y = \sum_{i=1}^{P} (Y_i - M_Y)^2 = Q_Y - P(M_Y)^2$$

The pooled estimate of the variance of the sample values and the group values, V_p, is given by

$$V_p = \frac{NSS_X + NSS_Y}{(N-1) + (P-1)}$$

A two-tailed t test is performed by comparing a calculated t value, t_{CAL}, against a critical value, t_{CRIT}.

$$t_{CAL} = \frac{M_X - M_Y}{\sqrt{V_p\left(\frac{1}{N} + \frac{1}{P}\right)}}$$

For a sample to be determined as statistically similar to a pixel group, the following relationship between the critical t value, t_{CRIT}, and t_{CAL} must be satisfied in each channel:

$$|t_{CAL}| < t_{CRIT}$$

The value, t_{CRIT}, is determined from a table of percentage points of the t distribution and is based on the confidence level chosen and the number of degrees of freedom, given by the quantity, (N-1) + (P-1).

In addition, the variance of the sample values, given by NSS_X/N, and the variance of the group values, NSS_Y/P, must satisfy the following:

$$\sqrt{NSS_X/N} < .15\ M_X$$

$$\sqrt{NSS_Y/P} < .15\ M_Y$$

This check on the variance requires that the standard deviation be less than 15% of the mean value in each channel. The constant value of .15 was derived empirically by LARS.

C. Second-Order Similarity Test

The second statistical test which has been employed by LARS to determine the similarity between a sample and a pixel group uses both first-order and second-order statistics. The test of the hypothesis that the means are equal is exactly the same as in the method just described. The difference results from the fact that this second method also tests the hypothesis that the variance of the sample is equal to the variance of the pixel group.

A two-tailed F test is performed by comparing a calculated F value,

F_{CAL}, against a critical F value for both tails of the F distribution, F_{CRITLO} and F_{CRITHI}, where F_{CAL} is given by the expression:

$$F_{CAL} = \frac{NSS_Y}{NSS_X} \quad \frac{N-1}{P-1}$$

The value of F_{CRIT} is determined from a table of percentage points of the F distribution and is based on the confidence level chosen, and the degrees of freedom, (N-1) and (P-1). For a sample to be assigned to a pixel group the following conditions must be satisfied in each channel:

$$|t_{CAL}| < t_{CRIT}$$

$$F_{CRITLO} < F_{CAL} < F_{CRITHI}$$

APPENDIX F

DERIVATION OF DISTRIBUTION OF BOUNDARY LOCATION ESTIMATE

The received signal $r(y)$ can be expressed in terms of the true boundary location γ_o as

$$r(y) = s(y-\gamma_o) + n(y)$$

The expression can be substituted into the partial derivative of the log-likelihood function to obtain the maximum likelihood estimate condition

$$\int_{-Y}^{Y} s(y-\gamma_o)\, s'(y-\hat{\gamma})dy + \int_{-Y}^{Y} n(y)s'(y-\hat{\gamma})dy - \int_{-Y}^{Y} s(y-\hat{\gamma})s'(y-\hat{\gamma})dy = 0$$

For a high signal-to-noise ratio, $\hat{\gamma}$ will be approximately equal to γ_o, the true value. The first term in the above expression can then be expanded in a Taylor series about the value $\hat{\gamma} = \gamma_o$. Making a change of variable

$$\int_{-Y}^{Y} s(y-\gamma_o)s'(y-\hat{\gamma})dy = \int_{-Y-\hat{\gamma}}^{Y-} s(y+\hat{\gamma}-\gamma_o)s'(y)dy$$

and using the first two terms of the Taylor expansion (additional terms are assumed neglibible), this quantity equals

$$\int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} s(y)s'(y)dy + (\hat{\gamma}-\gamma_o) \int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} [s'(y)]^2 dy$$

The original equation can now be written as

$$\int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} s(y)s'(y)dy + (\hat{\gamma}-\gamma_o) \int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} [s'(y)]^2 dy + \int_{-Y}^{Y} n(y)s'(y-\hat{\gamma})dy$$

$$- \int_{-Y}^{Y} s(y-\hat{\gamma})s'(y-\hat{\gamma})dy = 0$$

The first and last terms on the left-hand side cancel out and the quantity, $(\hat{\gamma}-\gamma_o)$, can be solved for.

$$\hat{\gamma} - \gamma_o = \frac{-\int_{-Y}^{Y} n(y)s'(y-\hat{\gamma})dy}{\int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} [s'(y)]^2 dy}$$

The estimate $\hat{\gamma}$ equals a constant plus a linear function of the noise $n(y)$. The noise is a Gaussian random process and, therefore, the estimate $\hat{\gamma}$ is a Gaussian random variable. The mean of $\hat{\gamma}$ is given by

$$E\{\hat{\gamma}\} = \gamma_o$$

The estimate is, therefore, unbiased.

The variance of $\hat{\gamma}$ is given by

$$V\{\hat{\gamma}\} = E\{(\hat{\gamma}-\gamma_o)^2\}$$

$$= \frac{E\{[\int_{-Y}^{Y} n(y)s'(y-\hat{\gamma})dy]^2\}}{[\int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} [s'(y)]^2 dy]^2}$$

The numerator of this expression is equivalent to

$$E\{\int_{-Y}^{Y}\int_{-Y}^{Y} n(u)n(v)s'(u-\hat{\gamma})s'(v-\hat{\gamma})dudv\}$$

which equals

$$\int_{-Y}^{Y}\int_{-Y}^{Y} R_n(u-v)s'(u-\hat{\gamma})s'(v-\hat{\gamma})dudv$$

Here, $R_n(u-v)$ is the correlation function of the noise. For white, Gaussian noise

$$R_n(y) = \frac{N_o}{2} \delta(y)$$

and the double integral is equal to

$$\frac{N_o}{2} \int_{-Y-\hat{\gamma}}^{Y-\hat{\gamma}} [s'(y)]^2 dy$$

Because $\hat{\gamma}$ is approximately equal to γ_o, the variance of $\hat{\gamma}$ can be written as

$$V(\hat{\gamma}) = \frac{N_o}{2 \int_{-Y-\gamma_o}^{Y-\gamma_o} [s'(y)]^2 dy}$$

For the forms of the signal s(y) which are of interest in this application, $s'(y)$ is equal to zero within a short distance of the point y=0. Provided γ_o is not too close to the endpoints of the interval, the denominator term in the estimate variance will be a constant and the variance can be written as

$$V(\hat{\gamma}) = \frac{N_o}{2 \int_{-Y}^{Y} [s'(y)]^2 dy}$$

The minimum variance of any estimate $\hat{\gamma}$ of the boundary location is given by the Cramer-Rao bound as

$$V_{min}(\hat{\gamma}) = \left[E \left\{ \left[\frac{\partial \Lambda}{\partial \gamma} \right]^2 \right\} \right]^{-1}$$

Substituting for the partial derivative,

$$V_{min}(\hat{\gamma}) = \left[E \left\{ \left[\frac{2}{N_o} \int_{-Y}^{Y} [r(y)-s(y-\gamma)]S'(y-\gamma)dy \right]^2 \right\} \right]^{-1}$$

Using the relation that

$$n(y) = r(y) - s(y-\gamma)$$

and the correlation function of the white noise, the minimum variance is given in the high signal-to-noise ratio case as

$$V_{min}(\hat{\gamma}) = \frac{N_o}{2\int_{-Y}^{Y} [S'(y)]^2 dy}$$

This is precisely the variance which has been derived for the maximum likelihood estimation procedure. Given the assumptions which have been made, no other estimation procedure will result in a lower variance for the estimated boundary location. The maximum likelihood procedure is optimal in the sense of minimum variance.

GLOSSARY

a — Letter denoting a material to be recognized by multispectral processing. It often refers to the material chosen by the decision rule.

AVE9 — A classification rule that averages the nine data vectors in the 3 x 3 pixel neighborhood and then applies the one-point rule QRULE to the result. To lessen sensitivity to alien points, the t largest and t smallest data values in each channel are deleted and the remaining values averaged. Sometimes called the "moving average" rule or the "trimmed mean" rule.

b — Letter denoting a material to be recognized by multispectral processing.

BAYES8 — A classification rule similar to BAYES9 except that the smallest square bracket factor in an old form of the decision criterion

$$P(X_o|a) \prod_{i=1}^{8} [P(X_i|a) + \frac{1-\theta}{\theta k} \sum_b P(X_i|b)]$$

is omitted

BAYES9 — A classification rule based on the assumption that a pixel probably represents the same material as its neighbor, the degree of dependence specified by a parameter θ between 0 (independence) and 1 (complete dependence). The decision criterion is

$$P(X_o|a) \prod_{i=1}^{8} [\frac{P(X_i|a) + S \sum_b P(X_i|b)}{\varepsilon + S \sum_b P(X_i|b)}]$$

criterion — An expression calculated by a classification rule for each material at every pixel. The material with the biggest criterion (sometimes the smallest) is the one chosen. A null test criterion is an expression which, when smaller (or sometimes larger) than a prescribed constant, signals a decision "none of these".

DATUM — A channel vector that is passed from module to module of the ERIM multispectral processing system. It starts out being the vector of multispectral data values at each pixel. After application of the one-point processing module QRULE, DATUM(1) is the QRULE choice, DATUM(2) the QRULE null criterion and DATUM(3)...DATUM(k+2) are the exponents of the multivariate normal density conditional on the materials being recognized (see "exponent".) After applying a nine-point rule, DATUM(1) is the material chosen and DATUM(2) the null criterion.

error rate — The error rate for one field is the percent misclassified. The error rate of a group of fields is the arithmetic average of the error rates of the fields.

exponent — The multivariate normal density can be written

$$\frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}[(X-u)^T R^{-1}(X-u) + \log_e |R|]}$$

where u is the mean vector and R the covariance matrix. The expression in square brackets is sometimes referred to in this report as the "exponent".

k — The number of materials to be recognized for which distributions have been specified.

LACIE — Large Area Crop Inventory Experiment, an experiment to estimate from satellite data the acreage of wheat grown in various wheat-producing countries.

LACIE intensive study site — A 3 x 3 mile area where LACIE data from the LANDSAT satellite is correlated with ground truth.

LANDSAT — The new name of the ERTS satellite that provides data for measuring earth resources.

level — The decision rules BAYES9, PRIOR9, and PREF9 decide null (i.e., "none of these") when the winning material density is less than a prescribed number ε. Level is the probability that a legitimate

point from a material distribution will be so rejected. The relation between ε and level is

$$\varepsilon = e^{-1/2\ (\mathrm{EXPLIM} + \log_e |R|)}$$

where $|R|$ is the determinant of the covariance matrix and EXPLIM is the value in the chi-square table corresponding to the row NCHAN (the number of channels of data) and the column "level".

LIKE9 — The maximum likelihood classification rule derived from the assumption that the nine pixels in the 3 x 3 pixel neighborhood are an independent random sample from a normal multivariate distribution. It amounts to adding for each material, the nine exponents (see "exponent") and then choosing the material with the smallest sum. To prevent occasional alien points from disturbing the decision rule, LIKE9 is modified to sum only the m smallest exponents, where m = 1,...9.

LOG2E — A table used in the rapid calculation of $\log_2$ as shown in Appendix A. It presents the result of subtracting a 3-octal-digit number from 200_8 and shifting left nine bits.

m — The LIKE9 criterion is the sum of the m smallest among the nine exponents. See LIKE9.

MAD — Michigan Algorithm Decoder is a computer source language resembling ALGOL.

nine-point rule — A nine-point classification rule decides what material to assign to a pixel on the basis of data from that pixel and its eight immediate neighbors.

null test — A provision of a classification rule to decide "none of these", i.e., to decide that a pixel represents none of the materials for which distributions have been specified. The null test used in BAYES9, PRIOR9 and PREF9 is based on the assumption that materials in the scene not given a specific distribution may be lumped together in a null category that has a flat distribution of height ε. When this distribution wins, a decision "none of these" is made.

P — "probability that". $P(X|a)$ means "probability of data vector X given that the pixel represents material a". $P(a|X)$ is "probability that the pixel represents material a given that the data vector is X."

p — probability that neighboring pixels represent the same material.

p_{ia} — The probability $P(a|X_i)$ that the i^{th} of the nine-pixels is a given that the prior probabilities are equal and the data vector is X_i. By Bayes' theorem it is

$$\frac{P(X_i|a)}{\varepsilon + \sum_b P(X_i|b)}$$

pixel — a resolution element of the multispectral scanning system. The continuous signal from a scan line is sampled to make it suitable for digital processing. The patch of ground corresponding to one sample is a pixel.

posterior probability — Given the prior probability $P(a)$ that a pixel represents material a and a known probability $P(X|a)$ of getting X as the data vector given that the pixel represents a, the posterior probability $P(a|X)$ that the pixel represents a given that the data vector is X is

$$P(a|X) = \frac{P(a)\ P(X|a)}{\sum_b P(b)\ P(X|b)}$$

by Bayes formula. The posterior probability of a is thus the prior probability modified by information from the data vector X.

PREF9 — A classification rule that uses the criterion "sum of the nine individual posterior probabilities", which is

$$\sum_{i=0}^{8} p_{ia} = \sum_{i=0}^{8} \frac{P(X_i|a)}{\varepsilon + \sum_b P(X_i|b)}$$

prior probability — The prior probability $P(b)$ that a pixel represents material b is the probability of b estimated without reference to the data from that pixel. In the absence of other information, the prior probabilities are often taken to be equal.

PRIOR9 — A classification rule embodying the principle that the prior likelihood that a pixel represents a given material is dependent on the neighborhood in which it is located. Specifically, the averaged posterior probability over the 3 x 3 pixel neighborhood becomes the prior probability for a Bayesian decision on the center pixel. The decision criterion is

$$p_{oa} \quad \sum_{i=0}^{8} \quad p_{ia}$$

and the null test is

$$\frac{p_{oa} \quad \sum_{i=0}^{8} p_{ia}}{\frac{1}{\varepsilon + \sum_{b} P(X_o|b)} + \ldots + \frac{1}{\varepsilon + \sum_{b} P(X_o|b)}} < \varepsilon_2^{\,2}$$

QRULE — The "quadratic rule", i.e., the usual, one-point maximum likelihood classification rule. It puts out the one-point decision as DATUM(1), the one-point null criterion as DATUM(2), and in use with nine-point rules, is run to put out an exponent for each material in DATUM(3),...,DATUM(k+2).

S — $\frac{1-\theta}{\theta(k+1)}$ where k is the number of materials and θ is the dependence parameter of BAYES9

SAVE9 — An array that saves two scan lines to provide the data for the 3 x 3 array used in nine-point rules. See Appendix B.

Signature — The mean and covariance matrix that specify the multivariate normal distribution of a material.

SIGS1 — A choice of 20 training fields from the 42 Imperial Valley field interiors used to test nine-point rules. Also refers to a choice of training fields in the proposed plan to test nine-point rules on satellite data.

SIGS2 — A choice of 20 training fields from the 22 Imperial Valley fields not in SIGS1. Also refers to a similar arrangement in the proposed plan to test nine-point rules on satellite data.

TALLY — A processing module that counts the number of recognitions of each material in a prescribed rectangle.

t — In the moving average rule AVE9, the t largest and t smallest data values in each channel are discarded or "trimmed" before the data values are averaged. This minimizes the disturbance of a noisy data vector.

X — The vector of data values at a pixel

VOTE9 — A classification rule, applied after one-point decisions have been made on the nine pixels, which assigns to the center pixel the material most frequently recognized in the nine pixels.

ε — A quantity used by the null tests for BAYES9, PRIOR9, and PREF9. All materials in the scene that don't have a specific distribution are lumped into a null category that has a flat distribution of height ε. The rules are applied as if this null distribution were a legitimate distribution when it wins, a decision "none of these" is made.

ε_2 — The BAYES9 null test, for example, is to decide null if

$$P(X_o|a) \prod_{i=1}^{8} \left[\frac{P(X_i|a) + S \sum_b P(X_i|b)}{\varepsilon + S \sum_b P(X_i|b)} \right] < \varepsilon_2$$

Theoretically, ε_2 is the same as ε, but because it can be reset after BAYES9 processing and ε cannot, it is convenient to refer to it by a different symbol.

θ — The dependence parameter of the classification rule BAYES9. $\theta = 0$ means neighboring pixels are completely independent. $\theta = 1$ means neighboring pixels represent the same material. θ between 0 and 1 means neighboring pixels probably represent the same material.

REFERENCES

1. W. Richardson, A Study of Some Nine-Element Decision Rules, Technical Report 190100-32-T, Environmental Research Institute of Michigan, Ann Arbor, Michigan, 1974.

2. R. F. Nalepka, Investigation of Multispectral Discrimination Techniques, Technical Report 2264, 12-F, Willow Run Laboratories, Ann Arbor, Michigan, 1970.

3. R. B. Crane and W. Richardson, Performance Evaluation of Multispectral Scanner Classification Techniques, Proceedings of the Eighth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, 1972.

4. R. B. Crane, W. Richardson, and W. A. Malila, A Study of Techniques for Processing Multispectral Data, Technical Report 31650-155-T, Environmental Research Institute of Michigan, Ann Arbor, 1973.

5. H. M. Horwitz, P. D. Hyde, W. Richardson, Improvements in Estimating Proportions of Objects from Multispectral Data, Technical Report 190100-25-T, Environmental Research Institute of Michigan, Ann Arbor, 1974.

6. R. Kettig & D. Landgrebe, "Automatic Boundary Finding and Sample Classification of Remotely Sensed Multispectral Data", LARS Information Note 041773, Purdue University.

7. J. N. Gupta & P. A. Wintz, "Closed Boundary Finding Feature Selection and Classification Approach to Multi-Image Modeling", LARS Information Note 062773, Purdue University.

8. P. Anuta, "Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques", IEEE Trans. on Geoscience Electronics, pp. 353-368, October 1970.

9. A. Rosenfeld, "Picturing Processing by Computer", Academic Press, New York, 1969.

10. R. Duda & P. Hart, "Pattern Classification and Scene Analysis", Wiley & Sons, New York, 1973.

Technical and Final Report Distribution List

NASA Contract NAS9-14123

Tasks II thru X

NAME	NUMBER OF COPIES
NASA/Johnson Space Center Earth Observations Division Houston, Texas 77058	
ATTN: Mr. Robert MacDonald/TF	(1)
ATTN: Mr. B. Erb/TF2	(1)
ATTN: Dr. F. Hall/TF2	(1)
ATTN: Mr. J. Murphy/TF2	(1)
ATTN: Dr. A. Potter/TF3	(8)
ATTN: Mr. J. Dragg/TF4	(1)
ATTN: Earth Resources Data Facility/TF12	(8)
NASA/Johnson Space Center Earth Resources Program Office Office of the Program Manager Houston, Texas 77058	
ATTN: Mr. Clifford E. Charlesworth/HA	(1)
ATTN: Mr. William E. Rice/HA	(1)
NASA/Johnson Space Center Earth Resources Program Office Program Analysis & Planning Office Houston, Texas 77058	
ATTN: Dr. O. Glenn Smith/HD	(1)
NASA/Johnson Space Center Earth Resources Program Office Systems Analysis and Integration Office Houston, Texas 77058	
ATTN: Mr. Richard A. Moke/HC	(1)
ATTN: Mr. M. Jay Harnage, Jr./HC	(1)
NASA/Johnson Space Center Technical Library Branch Houston, Texas 77058	
ATTN: Ms. Retha Shirkey/JM6	(4)

NAME	NUMBER OF COPIES
NASA/Johnson Space Center Management Services Division Houston, Texas 77058 ATTN: Mr. John T. Wheeler/AT3	(1)
NASA/Johnson Space Center Technical Support Procurement Houston, Texas 77058 ATTN: Mr. J. Haptonstall/BB63	(1)
Earth Resources Laboratory, GS Mississippi Test Facility Bay St. Louis, Mississippi 39520 ATTN: Mr. D. W. Mooneyhan	(1)
EROS Data Center U.S. Department of Interior Sioux Falls, South Dakota 57198 ATTN: Mr. G. Thorley	(1)
Department of Mathematics Texas A&M University College Station, Texas 77843 ATTN: Dr. Larry Guseman	(1)
NASA/Johnson Space Center Computation & Flight Support Houston, Texas 77058 ATTN: Mr. Eugene Davis/FA	(1)
U.S. Department of Agriculture Agricultural Research Service Washington, D.C. 20242 ATTN: Dr. Robert Miller	(1)
U.S. Department of Agriculture Soil & Water Conservation Research Division P.O. Box 267 Weslaco, Texas 78596 ATTN: Dr. Craig Wiegand	(1)

NAME	NUMBER OF COPIES
U.S. Department of Interior Geology Survey Washington, D.C. 20244 ATTN: Dr. James K. Anderson	(1)
Director, Remote Sensing Institute South Dakota State University Agriculture Engineering Building Brookings, South Dakota 57006 ATTN: Mr. Victor I. Myers	(1)
U.S. Department of Interior Fish & Wildlife Service Bureau of Sport Fisheries & Wildlife Northern Prairie Wildlife Research Center Jamestown, North Dakota 58401 ATTN: Mr. Harvey K. Nelson	(1)
U.S. Department of Agriculture Forest Service 240 W. Prospect Street Fort Collins, Colorado 80521 ATTN: Dr. Richard Driscoll	(1)
U.S. Department of Interior Geological Survey Water Resources Division 500 Zack Street Tampa, Florida 33602 ATTN: Mr. A.E. Coker	(1)
U.S. Department of Interior Director, EROS Program Washington, DC 20244 ATTN: Mr. J. M. Denoyer	(1)
U.S. Department of Interior Geological Survey GSA Building, Room 5213 Washington, DC 20242 ATTN: Mr. W.A. Fischer	(1)

NAME	NUMBER OF COPIES
NASA Wallops Wallops Station, Virginia 23337	
ATTN: Mr. James Bettle	(1)
Purdue University Purdue Industrial Research Park 1200 Potter Drive West Lafayette, Indiana 47906	
ATTN: Dr. David Landgrebe	(1)
ATTN: Dr. Philip Swain	(1)
ATTN: Mr. Terry Phillips	(1)
U.S. Department of Interior EROS Office Washington, DC 20242	
ATTN: Dr. Raymond W. Fary	(1)
U.S. Department of Interior Geological Survey 801 19th Street, N.W. Washington, DC 20242	
ATTN: Mr. Charles Withington	(1)
U.S. Department of Interior Geological Survey 801 19th Street, N.W. Washington, DC 20242	
ATTN: Mr. M. Deutsch	(1)
U.S. Geological Survey 801 19th Street, N.W., Room 1030 Washington, DC 20242	
ATTN: Dr. Jules D. Friedman	(1)
U.S. Department of Interior Geological Survey Federal Center Denver, Colorado 80225	
ATTN: Dr. Harry W. Smedes	(1)

NAME	NUMBER OF COPIES
U,S, Department of Interior Geological Survey Water Resources Division 901 S. Miami Ave. Miami, Florida 33130 ATTN: Mr. Aaron L. Higer	(1)
University of California School of Forestry Berkeley, California 94720 ATTN: Dr. Robert Colwell	(1)
School of Agriculture Range Management Oregon State University Corvallis, Oregon 97331 ATTN: Dr. Charles E. Poulton	(1)
U,S. Department of Interior EROS Office Washington, DC 20242 ATTN: Mr. William Hemphill	(1)
Chief of Technical Support Western Environmental Research Laboratories Environmental Protection Agency P,O, Box 15027 Las Vegas, Nevada 89114 ATTN: Mr. Leslie Dunn	(1)
NASA/Langley Research Mail Stop 470 Hampton, Virginia 23365 ATTN: Mr. William Howle	(1)
U,S. Geological Survey Branch of Regional Geophysics Denver Federal Center, Building 25 Denver, Colorado 80225 ATTN: Mr. Kenneth Watson	(1)

NAME	NUMBER OF COPIES
NAVOCEANO, Code 7001 Naval Research Laboratory Washington, DC 20390 ATTN: Mr. J. W. Sherman, III	(1)
U.S. Department of Agriculture Administrator Agricultural Stabilization and Conservation Service Washington, DC ATTN: Mr. Kenneth Frick	(1)
Pacific Southwest Forest & Range Experiment Station U.S. Forest Service P. O. Box 245 Berkeley, CA 94701 ATTN: Mr. R. C. Heller	(1)
United States Department of Agriculture/Forestry Service Division of Forest Economics and Marketing Resources 1200 Independence Avenue Washington, D.C. 20250 ATTN: Dr. P. Weber	(1)
University of Texas at Dallas Box 688 Richardson, Texas 75080 ATTN: Dr. Patrick L. Odell	(1)
Department of Mathematics University of Houston Houston, Texas 77004 ATTN: Dr. Henry Decell	(1)
Institute for Computer Services and Applications Rice University Houston, Texas 77001 ATTN: Dr. M. Stuart Lynn	(1)

NAME	NUMBER OF COPIES
U.S. National Park Service Western Regional Office 450 Golden Gate Avenue San Francisco, California 94102 ATTN: Mr. M. Kolipinski	(1)
U.S. Department of Agriculture Statistical Reporting Service Washington, DC 20250 ATTN: D. H. VonSteen/R. Allen	(2)
U.S. Department of Agriculture Statistical Reporting Service Washington, DC 20250 ATTN: Mr. H. L. Trelogan, Administrator	(1)
Department of Watershed Sciences Colorado State University Fort Collins, Colorado 80521 ATTEN: Dr. James A. Smith	(1)
Lockheed Electronics Co. 16811 El Camino Real Houston, Texas 77058 ATTN: Mr. R. Tekerud	(1)
TRW System Group Space Park Drive Houston, Texas 77058 ATTN: Dr. David Detchmendy	(1)
IBM Corporation 1322 Space Park Drive Houston, Texas 77058 ATTN: Dr. D. Ingram	(1)
S&D - DIR Marshall Space Flight Center Huntsville, Alabama 35812 ATTN: Mr. Cecil Messer	(1)

NAME	NUMBER OF COPIES
Code 168-427 Jet Propulsion Laboratory 4800 Oak Grove Drive Pasadena, California 91103	
ATTN: Mr. Fred Billingsley	(1)
NASA Headquarters Washington, DC 20546	
ATTN: Mr. W. Stoney/ER	(1)
ATTN: Mr. Leonard Jaffe/ER	(1)
ATTN: Mr. M. Molloy/ERR	(1)
ATTN: Mr. James R. Morrison	(1)
Ames Research Center National Aeronautics and Space Administration Moffett Field, California 94035	
ATTN: Dr. D. M. Deerwester	(1)
Goddard Space Flight Center National Aeronautics and Space Administration Greenbelt, Maryland 20771	
ATTN: Mr. W. Nordberg, 620	(1)
ATTN: Mr. W. Alford, 563	(1)
Lewis Research Center National Aeronautics and Space Administration 21000 Brookpark Road Cleveland, Ohio 44135	
ATTN: Dr. Herman Mark	(1)
John F. Kennedy Space Center National Aeronautics and Space Administration Kennedy Space Center, Florida 32899	
ATTN: Mr. S. Claybourne/FP	(1)
NASA/Langley Mail Stop 214 Hampton, Virginia 23665	
ATTN: Mr. James L. Raper	(1)

NAME	NUMBER OF COPIES
Texas A&M University Institute of Statistics College Station, TX 77843 ATTN: Dr. H. O. Hartley	(1)
Texas Tech University Department of Mathematics P. O. Box 4319 Lubbock, TX 79404 ATTN: Dr. T. Boullion	(1)
Mr. James D. Nichols Space Sciences Laboratory, Rm. 260 University of California Berkeley, CA 94720	(1)
EXXON Production Research Co. P. O. Box 2189 Houston, TX 77001 ATTN: Mr. J. O. Bennett	(1)